LECTURES ON THE SCIENTIFIC BASIS OF MEDICINE
1952–53

British Postgraduate Medical Federation
University of London

LECTURES ON THE SCIENTIFIC BASIS OF MEDICINE

Volume II

1952–53

UNIVERSITY OF LONDON
THE ATHLONE PRESS
1954

Published by
THE ATHLONE PRESS
at the Senate House, London, W.C.1
Distributed by Constable & Co. Ltd.
12 Orange Street, London, W.C.2

U.S.A.
John de Graff, Inc.
64 West 23rd Street,
New York, 10

Printed in Great Britain by
WESTERN PRINTING SERVICES LTD.
BRISTOL

PREFACE

With the publication of this the second annual volume of the series of lectures on 'The Scientific Basis of Medicine' and the preparation of the third annual volume now well advanced, it is possible to define more clearly the nature of the contribution which this series is designed to make to postgraduate medical education.

The British Postgraduate Medical Federation has entrusted to a committee of heads of departments of its component Institutes the task of arranging each year a series of thirty or more lectures on 'The Scientific Basis of Medicine' intended, as explained in the Preface to Volume I, for younger research workers and graduates undergoing training for careers as consultants and specialists in the various clinical branches of medicine and surgery. This committee reviews the fields of research in which there have been notable advances in the knowledge of fundamental principles that may later be applied to a better understanding of health and disease, to the improvement of health and to the prevention and treatment of disease. These advances may be in the fields of physics, chemistry, anatomy, physiology, biochemistry, pharmacology, pathology, bacteriology, immunology or clinical science; and the committee takes account not only of advances the application of which to the practice of medicine is already evident but also others about the future application of which it may at the given time be possible to form no more than a conjecture. In the light of this review a selection is made of topics which seem to the committee ripe for presentation in lecture form to an audience of young medical research workers and junior consultants and specialists. The lecturer invited to deal with each selected topic is a person actively engaged in furthering advance in it by original research.

In some fields advance may be proceeding over a wide front,

and it may be appropriate to devote two or more lectures to it in the same series, and to cover several aspects; in others a single lecture may suffice to present those developments that are at once recent and significant. In either case each subject is kept under constant review, and as further advances are made, further treatment is given to it in later lecture-series. In this way the Federation seeks to ensure, on the one hand, that every field of research which develops a fresh importance for medicine is included in the series as soon as its findings reach the stage at which they can be reviewed in lecture form and, on the other, that the developments stimulated by every new advance find their due place in the course over a period of years. To give a few concrete examples: the discovery of radioactive isotopes, and the subsequent advances in their preparation and in the techniques for their use in biological and clinical research have opened up new fields of investigation into the chemistry of living matter, into the problems of the nutrition of body cells and their life span, into the actions of hormones and enzymes, and into the physiology of endocrine glands and their interrelations. The discovery of a clinical treatment such as that by penicillin in certain infections gives clues for fresh investigation in bacteriology, immunology, virology and chemotherapy. Or again, the discovery of the beneficial action of ACTH in rheumatic diseases has stimulated a renewal of interest in the connective tissues and body fluids and in the physiology of the pituitary and adrenal glands.

The lecture syllabuses are not, therefore, mere haphazard tours around a changing scene, nor can any single annual syllabus properly be treated as an isolated entity. Each course of lectures is planned as a stage in an integrated, continuously developing survey, which can only yield its full value to the student who keeps himself continuously familiar with it.

It is in order to enable students to do this that the Federation has decided to provide a permanent record of each lecture course in the form of an annual published volume. It is not practicable, and would not be appropriate, to publish the whole group of lectures delivered each year. Some topics can be usefully presented in lecture form before they are ready for evalu-

ation in print; for some others an accessible summary may already exist. In cases of the former kind the subject will be taken up again when an authoritative assessment of its main advances can be made. It is the hope of the Federation that each volume fairly reflects the main contents of the particular group of lectures on which it is founded; and that, taken together as a series over a period of years, the volumes will be found to offer a reliable guide to all the main advances in knowledge that derive from research in the basic sciences, and have or are likely to have clinical applications. Amongst the articles in the present volume, for instance, will be found not only articles concerned with antibiotics and ACTH, and the new researches developing from discoveries in these fields, but also a number dealing with the chemical transmission of nervous impulses, the preservation of living cells at low temperatures, and the use of radio-isotopes and chromatography. Further developments of these themes will be found in Volume III, which should be ready for publication in the spring of 1955; and it is hoped that many of the important contributions now being made by physicists to the understanding of biological problems will be reflected in the volumes to be published in 1956 and later years.

FRANCIS R. FRASER
Director, British Postgraduate Medical Federation

3 February 1954

NOTE

The lectures printed in this volume were delivered on the following dates:

I.	6 January 1953	X.	15 January 1953
II.	25 and 27 November 1952	XI.	11 December 1952
III.	29 January 1953	XII.	9 December 1952
IV.	4 December 1952	XIII.	13 January 1953
V.	2 December 1952	XIV.	20 January 1953
VI.	26 February 1953	XV.	24 February 1953
VII.	3 February 1953	XVI.	17 February 1953
VIII	16 October 1952	XVII.	12 February 1953
IX.	20 November 1952	XVIII.	6 November 1952

XIX. 14 October 1952

CONTENTS

I

The Methodology of Clinical Science

SIR JAMES SPENCE

THERE is a risk in generalizing from the few facts which we apprehend in our personal experience, and I take this risk in saying that there are two main kinds of clinical science, of which one is in danger of being overwhelmed by the other. Out of this distinction I shall try to create the substance of this lecture.

Clinical science is the planned study of disease, or of the phenomena of disease, in men and women and children who are sick. It is a practical craft, and this implies that the clinical scientist must possess the particular skills necessary for collecting and systematizing his knowledge. He must possess also the opportunity to do his work with such regularity and method as the occasion demands Working within his responsibility for the medical care of the patient he must do nothing which is harmful to his patient or which delays his cure. If his researches lead the clinical scientist to do more than this or less than this, his intentions must be approved by the patient after full explanation of what is in store for him. Approval by the patient will always be influenced by the attitude and the opinion of the investigator, and there can be no tribunal to pass judgement on what is practice and what is malpractice in the investigation of sick patients. Under these circumstances research methods can be sanctioned only by that faculty which distinguishes right from wrong, and which we may still call conscience. This understanding of what is right and what is wrong in the care of patients is part of the professional knowledge gained in medical schools and teaching hospitals from teachers, who hold their teaching posts for the

has to be found, it would be better to call it 'clinical phenomenology'. Thomas Lewis, however, by the mere force of his personality, has so influenced clinical research in this country in recent years that the study of clinical phenomena, or 'states of disease' as he called them, overshadows other kinds of clinical [illegible]

Clinical Observation

Before I go further I must try to make clear the different kinds of clinical observation. That the instruments of clinical observation are now so disconcertingly numerous makes this all the more necessary. We should keep in mind, however, that it is the essence of clinical research that while the observer should know what to look for, he should not be too absorbed to notice the unusual. To grasp the significance of the unusual is an uncommon quality, and Walshe stressed this idea in saying 'it takes the exceptional mind to pass beyond the first exercises of the discriminatory tendency to pick up a hitherto unrecorded phenomenon, and to grasp its significance. Indeed many a competent clinician may live out his professional existence never having made a truly original observation.' With this chastening thought in our minds, we can try to separate clinical observation into its parts.

There are three main types of clinical observation, and each is as much an expression of a natural cast of mind, as of a trained method of approach; which is not to say that they cannot all be combined occasionally in one person's skill. How occasionally I cannot estimate, for I cannot recollect having seen such a person.

The first type of observation is the episodic recognition of spontaneously occurring signs of disease which we use when we diagnose [illegible] quickly [illegible] out of [illegible] degree. It comes best to those who talk least, and the ebullient, loquacious teacher rarely possesses it. It sees the rose spot and

very reason that they can transmit this knowledge, with other knowledge, to their assistants and students.

I return to the two kinds of clinical science. One way of approach to definitions is to look at their aims. The aim of one kind of clinical science is to *know* disease in such a way as to enable us to infer its source and to predict its course. Seen in this way, a disease is regarded as that sequence of phenomena in a sick person which arises from and follows on a noxious injury; and this is in accord with the Aristotelian idea of the nature of things. The methodology of this kind of clinical science is the planned observation of the sequence of phenomena in that number of sick people which allows us to know the common course of the disease under study from its beginning to its end, and to estimate and know the variations from that course. Planned observation implies the use of all practicable and necessary machines of measurement. The knowledge it leads to has immense practical values, for to know disease in this way is the basis of exact therapy in professional practice. This kind of work might be called clinical cartography, because it helps the practitioner to know where he is, and where he should go

The other kind of clinical science might be called clinical phenomenology. It studies the isolated phenomena of disease or of disordered function, and its aim is to explain their mechanisms and their clinical significance. It is concerned with auricular fibrillation, hunger pain, itchy skin or hyperpotassemia, rather than with the diseases in which these phenomena arise. Its methodology starts with identification of the phenomenon and then goes on to the use of experiment when possible and necessary. Great benefits have come to both the science and practice of medicine by these detailed analytical studies of clinical phenomena The knowledge gained has increased our understanding of old diseases and pointed the way to the discovery of new diseases.

Thomas Lewis did so much to establish this precise and rational study of clinical phenomena, that it is tempting to use his name as an eponymous title for this kind of clinical research, but Lewis' view in medicine went beyond detailed studies of phenomena and for that reason I have suggested that, if a name

artist's power to see something fresh in 'things we have passed a thousand times nor cared to see', but it has been an attribute also of most of the great scientists. To describe it I will mention the example only of N. M. Gregg's brilliant recognition of the association between rubella in the pregnant woman at her eighth to tenth week of pregnancy, and the congenital deformities in the child who will be born seven months later. If Gregg had not systematized his observations quickly at that time, and in those circumstances, the chance might not have come again, or for another hundred years or more. It was fortunate that the chance came to the sensitive mind. This kind of clinical observation, for which I can find no suitable name, is one of the most scintillating starting points of scientific research. But we cannot order it or command it. The best we can do is to recognize its possibility in the sensitive few, and to prevent its inhibition by too much teaching, its submersion by too much dogma, and its extinction by too much ritual, even though it be the ritual of research.

The third kind of clinical observation is what we seek to cultivate in clinical research. It is clinical observation planned to answer a carefully prepared question. The plan is controlled by well-known criteria: (*a*) that the question is worth answering; (*b*) that it has not already been answered by someone else; (*c*) that a plan of observation can be designed which is likely to provide an answer to the question and (*d*) that the observer is the sort of man who will carry out the plan. This kind of planned observation is one of the established methods of clinical research, and its scientific value is judged by the design of the plan Its purpose is to discover something, or to clarify some already existing field of knowledge. In clinical medicine the latter is not less important than the former, for growth of knowledge is not inevitable, and there is, I think, such a thing as a constant growth of ignorance in medicine, which can be kept in check only by opposing to it the clarifying process of research into the validity of existing knowledge.

Examples of planned clinical observation appear in various forms, but I choose the investigation of the outbreak, in 1948–9, of poliomyelitis in Canadian Esquimaux, at the Chesterfield

the scabies burrow. It hears the cry indicating a retropharyngeal abscess and the wheeze of tracheal pressure. It knows when to dismiss the record of a high blood urea as a laboratory mistake, or as the error of collecting the blood into a contaminated tube. In its best form it works towards an economy of effort, and has its eye on what is true and what is false in new machines. To this practical skill of the experienced clinical observer it is customary to give the name of clinical intuition; but if this implies a kind of mystical guessing, it no more deserves the name than does the skill of knowing where to find a plover's nest. It is the product not of guessing, but of a sifted experience by which the significant is recognized with such rapidity that the steps of reasoning are not discernible to the uninitiated. There is of course, and all too commonly, such a thing as clinical guessing, which we may suspect when we hear it bolstered up by expressions of opinion beginning with the words 'In my experience . . .', but this does not come into any of the categories of serious clinical observations we are discussing.

This episodic or incidental type of clinical observation is not, in itself, a method of research. It is the tool of professional practice. However skilfully it may be used, it is not likely to lead either to discovery or to a decisive clarification of existing knowledge, both of which are the objects of research. The truth of this will be known to those who have used the routine clinical notes of hospitals in seeking recorded facts and observations. These records may have their value as a teaching exercise, but as instruments of clinical research the routine ward notes and the machinery of hospital records are almost useless.

The second type of clinical observation is so rare that I may not make it evident except in examples. It comes only to the mind prepared It is Coleridge's 'ascertaining vision' which comes to a few 'who even in the level streams have detected elements, which neither the vale itself nor the surrounding mountains contained or could supply'. Its essence is the recognition of a *new phenomenon or of a new correlation between two* previously recognized phenomena Coming to those with a sufficient technical experience, it takes shape only in the minds of those endowed with imagination and insight. It is near to the

Planned Observation of Disease

In describing planned observation of a disease there is a danger that I create for it an artificial category. This danger is overcome if we assume that in some circumstances the researcher may use more than one trained skill. He may, for example, be both a trained clinical observer and a biochemist, or a morbid anatomist, but the combination rarely exists. Precise clinical observation of a disease, carried out to a pre-arranged plan, takes so much time and training that the observer has little chance of being competent in a second technical field.

Planned observation of a disease is simplified if the noxious agent and the time of the injury are known, but it is not the role of the clinical scientist to study the nature of the noxious agent. That work now properly devolves upon the microbiologist, the chemist, the physicist and the pharmacologist. Nor is it the role of the clinical scientist to be morbid anatomist or experimental pathologist, who see disease not as the clinical scientist sees it. Their concepts of disease are essential to the progress of knowledge, and medicine could neither exist nor advance without them, but it falls to the clinical scientist to see disease as the sequence of phenomena in the living which follows consequentially from an injury, and in identifying and measuring the phenomenal results to create a corpus of exact knowledge which can be applied in the practice of medicine. He maps the course of the disease by this sequence of phenomena subjectively interpreted by the patient and objectively identified by himself. His main task is to place the phenomena in temporal and in quantitative relationships with each other. This leads him to know the course of a disease as it may be expected commonly to occur. His next task is to determine the variations from that course, and to find correlations between these variations and aetiological factors or alternative treatments. When possible he uses statistics to express these variations He uses statistical estimates of variations also in designing the extent of his study. If the disease under study is one which varies little in its course, he limits his number of examples. If the course of the disease has many variations, or a wide 'lunatic fringe' as the clinical

Inlet, because it shows how observation may on occasions rise superior to the experimental method. If Peart, Rhodes, and their colleagues had not taken the opportunity to plan and carry out this observation the circumstances might not have occurred again, nor could anyone have designed an experiment to repeat the circumstances. Its significance lay in the fact that it was an outbreak of poliomyelitis in a virgin soil community, the like of which we are not likely to see again with the advent of the aeroplane. The picture was unique. It showed a case incidence of paralysis of 185 per 1,000 of the population; 1 in 20 of the population died; no children under the age of four years were paralysed; the epidemic reached its height in February with the mean temperature at −31° Fahrenheit. Here was *something unparalleled in the recorded epidemiology of the disease*, and the success of the observations depended on the team of five medical men who flew from Toronto to the scene of the epidemic, and carried their plans into effect. There are many other examples of planned clinical observation, in more obvious fields of enquiry, such as controlled therapeutic trials and studies in prognosis, but the essential features of all are the choice of a *question worth answering, the definition* and limitation of what is to be observed, the training and character of those who will make the observations, and the arrangements by which the material for observation is collected, or by which observers go to the material to make their observations.

Lister's planned observation of the strawberry naevus in 1938 will serve as a very straightforward example of design and accuracy. The opening words of his simple but near-perfect paper are worth noting:

> The idea of this investigation arose at a meeting of the Dermatological Section of the Royal Society of Medicine, when Mr Banks showed an infant with a large number of naevi. I was struck by the uncertainty of all those present as to the best line of treatment to be adopted, and as to what would happen if nothing was done.

He thereupon designed his observation, which lasted five years and comprised 77 patients, by which our knowledge of the natural history and treatment of the naevus has been established.

of the bases on which scientific clinical medicine is built. It may not reveal and unfold any new natural laws, but the practice of rational medicine relies on it, for without the power to predict the natural course of a disease, advice and treatment will remain at the mercy of insubstantial opinions.

My claim is that clinical science should correct the bias which has been imposed upon it by dominance of the experimental study of phenomena, and a few young clinical scientists should be trained in the exact observation and measurement of disease. If it be asked what value can these have, since diagnosis and treatment are the end-all and be-all of practical medicine, my reply would be that it would bring mature and trained observers to the bedsides of patients and that in those circum- [illegible] red. [illegible] east [illegible] for the pattern of these, changing as they do in the new environments which man creates for himself, are not static and eternal.

The inaccuracy and inadequacy of our recorded knowledge about disease and illness will become evident if we consult a few examples in text-books and journals. Nowhere in medical literature can I find a complete coherent account of the sequence of clinical events which follows the burning of a man, a woman or a child. If convulsions occur after burning is their frequency related to the age of the patient or to the extent of the burn? Is the time incidence of the convulsion, if it does occur, limited to the second day of the illness? There is no recorded answer to these questions. Many studies of isolated phenomena of burns have been successfully pursued, but the close relationship of one phenomenon to another has not been established That this should be the case in a disease so easy to observe and to measure is an index both of our ignorance and of faulty observation.

On the other hand, when we hear a young clinician say, after examining and observing a suspected case of meningococcal septicaemia, 'I have been watching the appearance and measur- [illegible]

scientist may call it in his despondent mood, he takes this into consideration in making his estimate of the number of patients to be taken into his study, and then proceeds towards that goal. If he takes less than he needs, his study is deceptively incomplete. If he takes more than he needs, he wastes his time. In this way he plans his study of disease, and with one research finished he passes on to study another disease. He thus comes to know disease as a predictable sequence of events, and the knowledge gained becomes the basis, the only basis, by which the underlying processes of disease in the living patient can be rationally interpreted.

I may be setting my aim too high in describing this method of study, or this method of study may have less scientific value than I think it has, but assuming that it is a necessary and valuable instrument for one kind of clinical research I believe it to be greatly neglected. Apart from a few elementary observations on pulse and temperature, how little we know about the detailed course of disease. How rarely does a clinical investigator sit through an illness recording the observations and instrumental measurements for himself. How seldom does the young physician sit through a night recording the pattern of behaviour, of delirium or of sleep in a patient with encephalitis. How infrequently does the young surgeon train himself in diagnosis by sitting at the bedside of a man with ruptured duodenal ulcer to measure the fluctuation of his signs and symptoms during the period while the operation is prepared. If the illness be prolonged how rarely do we design our observations with a regularity of twelve-hourly or even daily intervals. Most of us come to the patient to observe only the phenomena which custom has predetermined for us, or we take the observations secondhand from others who have recorded them for us. I am conscious of this in confessing that I, myself, know very few diseases in this discriminating manner which satisfies my scientific conscience. For example, I could not, to my own satisfaction, describe measles, or leukaemia, or acute rheumatism, or primary herpetiform stomatitis with their hour-to-hour, or day-to-day sequence of significant phenomena precisely determined, measured and accurately recorded. Yet this kind of knowledge is one

those whose skill is mainly in that branch of medicine. He runs into the danger of confusing clinical science with something called 'clinical investigation', and of regarding scientific method as synonymous with the use of diagnostic machinery. These misleading impressions are intensified by his preparation for final examinations with their diagnostic exercises as 'long cases' and 'short cases'.

These impressions about the attitude towards clinical research are strengthened by my experience on a Regional Hospital Board Committee which reviews applications for research grants from members of the medical staffs of the hospitals. The applications come almost entirely from members of the clinical staffs, and they fall roughly into four categories. The majority of applications are for the employment of morbid anatomists, or biochemists or technicians to do some laboratory research which the clinicians wish to have carried out, but in the techniques of which they themselves have had no laboratory training. The second category consists of requests for secretarial help in retrospective studies of hospital records. The third category is the request for diagnostic machinery. The smallest category is the request for the assistance of a clinical research worker to enlarge and complete a piece of clinical research which is already under way, and which the applicant is conducting for himself.

With moneys now available from the Universities, from endowment funds of hospitals, from the Government Research Councils and from the Educational Trusts the opportunities for clinical research are now beyond the dreams of what might have been considered possible thirty years ago. We can therefore look forward with great interest to the directions in which these forces will flow. Human nature being what it is, and human institutions what they are, we must expect that the opportunities will attract those who regard the publication of a research paper as a kind of higher diploma which will increase the chance of their being promoted to a higher clinical post. Others will use the opportunities to get from these resources the extra help they need for their routine clinical work in hospitals. These frailties can be easily humoured and helped in other ways

spots of meningococcal septicaemia are not that shape at this hour of the illness . . .', when we hear him describing his observations in that way, we know that he is heading in the direction of being a clinical scientist. But there are many hindrances to this training and many obstacles to that kind of clinical work, which I must now attempt to discuss.

HINDRANCES

This talk about scientists may build them up too much. No sensible young man starts out to be a clinical scientist. By accident or opportunity he later finds himself heading in that direction and follows the stream; or he comes under the personal influence of someone he likes and wishes to imitate; or he is just that sort of chap who must do research because he has a passion for getting things clear in his own mind, or for seeing how things tick over; and finding no other outlet or way of doing it he tries his hand at clinical research. He equips himself to become a clinical scientist, and then he is between the physical scientists who do not understand what he does, and the practitioners of medicine and surgery who try to understand what he does. He seeks a job either as a full-time professional research worker, or by combining research with some routine clinical practice The latter is the more difficult path to follow, particularly in that clamouring kind of hospital which is lively with many conferences and other distractions But he may order his life successfully by arranging clearly demarcated periods or days, in the way that Victor Horsley did, when he collects and systematizes his observations.

Hindrance to clinical research lies less in the claims of other clinical responsibilities than in the prevailing attitude of mind towards it, or the climate of opinion as it is now called. This attitude of mind begins to take shape in medical schools where the undergraduate thinks of medical research only as an affair of animal experiment and laboratories, but I must exclude from this generalization the three, or perhaps four *alpha* schools in this country which present a more wholesome picture of research. In the teaching hospitals he is immediately engaged in the excitements of diagnostic medicine, and his teachers are

tions beyond the hospitals, and extend these researches which can best be done in family practice and by field survey of random samples of population. The methodology of planned clinical observation in family practice and in random samples of the population will become a responsibility of medical schools. I foresee the day when medical faculties of Universities will have three provinces of clinical training and clinical research. One will be in hospitals, where training and research will be dependent on the facilities which remain after the demand for specialist forms of treatment have been satisfied. The second province will be a sample of population in their own homes which can be provided in family practice. The third will be in expeditions to communities, whether they be in a mining village, a Hebridean island, a Chesterfield Inlet, or wherever the enquiry takes them. If we are not too pompous about it all, if we can escape the dead hand of dull conformity, if we can get the freedom to make experiments in arranging our institutions, if we can recover the initiative to make the experiments, our medical schools and teaching hospitals may take on this work. If they fail it is difficult to see how else it can be done.

than by calling them research, but if they must be called research then we must find another name for the kind of work a man does when he devotes himself to that planned and patient observation of disease which is guided by a substantial working hypothesis.

Behind all these opportunities and arrangements I see the urgent need for a due proportion of young doctors who will become physicians and surgeons and family practitioners, to be trained in this kind of work, otherwise we shall face the position of which Adrian warned us in his introductory lecture to this course last year. 'The pathologists and biochemists will find', he said, 'that their time is taken up with measurements of uncertain value in which they are not specially interested, and the final result may well be that the work is turned over to specially trained technical experts who are the last people to give a dispassionate judgement on the value of what they are doing. . . .' He went on to indicate the cure for this in saying, 'The point, surely, is that if such data are to be valuable there must be the right people to consider them'

We are now concerned with the selection, the training and the working conditions of these right people. This brings me back to where I started. We are not likely to lack a sufficient number of the right people trained in the study of isolated clinical phenomena and their mechanisms. The influence of the schools of physiology and experimental pathology will continue to work in that direction. But I greatly fear that the very complexity of these phenomena is withdrawing interest from the planned study of their causative diseases. To that extent there is a growth of ignorance about these diseases. The tempo of events in some diseases may be so slow that they will be accurately observed only by a general practitioner, or by a family doctor over a period of many years; or the frequency of the disease may be so rare and its distribution so wide, that special arrangements may be required to gather a sufficient number of examples within the reach of the observers Hospitals, as we now know them, will not be the best places from which all this work can be organized. If clinical research is to be used to get a full picture of disease it must equip itself to carry the observa-

(Albert, 1951). Moreover a good deal of fresh light can be thrown on the subject by consideration of even remoter terri-

seem to be two departments of knowledge remote from the medical field. Yet the selectivity that enables Dispersol Yellow to dye exclusively the cellulose acetate in a fabric that contains both this fibre and plain cellulose (i.e. cotton) has something of importance to teach us about how a hypnotic accumulates selectively in the central nervous system.

The Practical Accomplishments of Selective Toxicity

The use of DDT in January 1944 to prevent an outbreak of typhus overrunning Europe was an historical event. The entire civil population of Naples was dusted with this substance to kill lice which were acting as vectors of the dreaded disease. It has been claimed that conditions at that time were such that the number of deaths from typhus would have exceeded the total casualties of the war. DDT is also effective for suppressing the breeding of mosquitoes and hence lowering the incidence of malaria. Much use of this substance is also made against household pests. Other selective agents used against ectoparasites include pyrethrum (flies), dimethyl phthalate (a repellent for adult mosquitoes), dibutyl phthalate (mites carrying scrub-typhus) and benzyl benzoate (scabies)

Turning now to endoparasites, we encounter that special branch of selective toxicity for which Ehrlich coined the name 'chemotherapy' at the beginning of the present century. Chemotherapy, as defined by Ehrlich, is the cure of an infectious disease using substances of *low* molecular weight (consequently sera and vaccines are excluded). Before Ehrlich's time the only known chemotherapeutic agents were quinine for malaria and crude ipecacuanha for amoebic dysentery The situation is now very different. Penicillin is used to cure syphilis, gonorrhoea, pneumonia and streptococcal septicaemia; chloramphenicol and aureomycin for typhus, sulphamezathine for meningococcal meningitis; sulphaguanidine and sulphathalidine for bacterial

II

Selective Toxicity

ADRIEN ALBERT

Fundamentals of Selective Toxicity

By Selective Toxicity we mean the poisoning of living cells without injury to other kinds of cells with which they are in contact. The cells which we wish to injure are those of the *uneconomic* species, and the cells which we wish to preserve are those of the *economic* species. These species may be related to one another as parasite and host, but in other cases they may be entirely symbiotic.

Before passing to the medical applications of selective toxicity, I should like to point out the immense importance of this subject in agriculture. The seeds of economically important plants, the soil around the plants and the plants themselves are normally infected with fungi, weeds and insects of all kinds. These threaten the economic success of the crops and a great deal of research has been carried out to discover substances which will destroy the pests without damaging the economic species. At the present time a large section of the chemical industry is concerned with the manufacture of these chemicals.

In animal husbandry, a similar situation exists. Man's economic animals are afflicted by all sorts of internal and external parasites, and many selectively toxic drugs have been devised to eliminate the invaders.

As I am addressing a medical audience, I propose to confine the remainder of these lectures to a discussion of selective toxicity in man. But in doing so I would emphasize that the application of this subject to man is constantly being enriched and illuminated by comparable studies of animals and plants

lower the patient's temperature, to stimulate (or depress) his sympathetic or parasympathetic nervous systems, to raise (or lower) his basal metabolic rate and to increase (or decrease) the clotting power of his blood. Deficiency or hyperactivity in muscle-action (including that of the heart) is subject to a certain amount of control as are the activities of a few of the endocrine glands. Excessive secretion of histamine, the cause of symptoms in hay-fever and urticaria, can now be counteracted. Cancer remains a baffling problem: prostatic cancer is readily controlled with oestrogens, and for some of the other forms of cancer there exist life-prolonging (but not life-saving) remedies. The immense amount of research now being done in this field should yield a richer harvest in the near future.

The Scientific Basis of Selective Toxicity

Selectivity, in some cases of toxic action, depends on nothing more than differences in accumulation. For example, every farmer knows that he can rid a wheatfield of chickweed by spraying with 20 per cent sulphuric acid. This acid is equally injurious to the cell-contents of both species, but it does not harm the wheat plant because this is smooth and waxy and hence sheds the poison, whereas the chickweed is rough and absorbent and accumulates the acid. In other cases, accumulation does not provide the explanation. For example, the powerful oestrogenic drug 'Fenocyclin' is concentrated 7 and 2 5 times as much in the intestines and liver respectively as it is in the uterus, and yet it acts only on the womb. This and many other examples indicate that selectivity can be brought about by a specific reaction, that is to say, one that takes place to a much greater extent in the uneconomic, as opposed to the economic, cells.

It is true that a common biochemical pattern runs through all living cells. For example, the Meyerhof sequence of glycolysis has been demonstrated in representative bacteria, protozoa, nematode worms, insects, echinoderms and vertebrates as well as in yeast and the higher plants. This process of glycolysis is found always to obtain its driving energy from phosphorylations by adenosine triphosphate. Moreover, many hormones have been shown to affect a wide range of species. For example,

dysentery; emetine, carbarsone, chiniofon and chloroquin for amoebic dysentery; atebrin, pentaquin, proguanil, chloranil and 'Daraprim' for malaria. Millions of people in warm, damp climates are cured of hookworm by carbon tetrachloride. Two other dread tropical diseases, schistosomiasis and filariasis, are beginning to yield to heterocyclic chemicals. Therapeutic help of some magnitude can be given in tuberculosis by *p*-aminosalicylic acid, streptomycin and isoniazide.

In spite of these successes, the field is still urgently in need of further exploration. Organisms tend to become resistant to frequently used drugs. Hence alternative substances, working through entirely different biochemical mechanisms, must be found. There is still no chemotherapeutic agent for hydatid infections, nor can any known drug cope with illness due to small viruses such as infantile paralysis and influenza.

So far we have dealt only with cases where the economic and the uneconomic cells are those of distinct species. We will now examine the more complex case where the uneconomic cells are part of the organism of the economic species. Thus the uneconomic cells may be a malignant growth in an otherwise healthy host, or they may be an overactive gland or some hyperfunctioning part of the nervous system. This branch of selective toxicity may be spoken of as pharmacology proper (but it will be kept in mind that pharmacology, in its wider aspects, covers drug studies of all kinds). In this field, selectively toxic agents are usually required to have reversible action. For example, the success of a patient's operation may depend on the use of an anaesthetic and a muscle-relaxing drug, but it would not be considered a success if these anaesthetic and relaxant effects should prove to be permanent.

In spite of the obvious difficulties of finding pharmacological agents with these desirable properties, many notable advances have been made in this field and remarkable control can now be exerted over the working of selected parts of the human body. Patients can be relieved of pain of all types and degrees of severity, put to sleep or (chemically) awakened, prevented from having convulsions or else caused to have them for therapeutic reasons. Other selectively toxic agents can be used to raise or

substances exhibiting an adrenaline-like action on the frog's heart and rabbit's ear have been found in protozoa, annelid worms, molluscs and arthropods (conversely, true adrenaline causes increased muscle-tone in annelida, molluscs and arthropods). Again, many vitamins, notably riboflavine, thiamine and nicotinic acid, are essential constituents of all living cells.

Hence it might seem at first sight as though chemical differences would not exist, of a magnitude sufficient to permit a toxic agent to injure one species and to leave another species unharmed. Closer investigation shows that this fear is unfounded. Glycolysis, it is true, follows the well-known course in very many species. However some alternative, apparently more primitive, mechanisms are also known to exist and these tend to predominate in bacteria which are some of the commonest uneconomic cells with which a physician has to deal. For example, bacteria often oxidize glucose or glucose-6-phosphate directly without preliminary conversion to phospho-glyceraldehyde. Moreover the tricarboxylic acid cycle, in which the raw materials of carbohydrate-breakdown are commonly synthesized into fat and protein, is imperfect in bacteria. Furthermore, adenosine phosphate, although universal in its distribution, functions in conjunction with creatine phosphate in vertebrates, and with arginine phosphate in molluscs, arthropods and worms. Neither of these two phosphagens are present in still lower forms of life.

This evidence of species difference in carbohydrate metabolism gives us some inkling of how the biochemistry of different kinds of cells can show individuality. Even greater differences are found in nitrogen metabolism. The end-products of this branch of metabolism range in complexity from the simple four-atom molecule of ammonia to the vast complexity of the alkaloids. Even among vertebrates, the end-products of protein and purine metabolism differ widely, as Table 1 shows, and they differ no less among the humbler forms of life.

Even more remarkable is the fact that the end-products of nitrogen metabolism can vary during the development of a *single species. For example, the chicken embryo excretes its* nitrogen mainly as ammonia for the first four days after fertilization, just as though it were a fish. But nine days after fertilization

it is excreting nitrogen mainly in the form of urea (more typical of amphibia) and only on the eleventh day does excretion occur mainly in the form of uric acid, which is the form typical of the adult. Thus it is understandable why the same selectively toxic agent is not suitable for a given parasite at all stages of development. For example, sexual and non-sexual forms of plasmodia coexist in the blood-stream of malarial patients. The sexual forms can be eliminated by plasmoquine or pentaquin but not by atebrin or proguanil, whereas the reverse is true for the non-sexual forms.

TABLE 1. End-products of Purine and Protein Metabolisms in various Phyla (Baldwin, 1948)

Phyla	Protein Metabolism	Purine Metabolism
Fish (teleosts)	ammonia	urea
Amphibia	urea	urea
Reptiles (snakes)	uric acid	uric acid
Birds	uric acid	uric acid
Mammals	urea	allantoin*

* Excepting man, higher apes and the Dalmatian dog which excrete uric acid.

It was pointed out above that at least three substances that are vitamins for man are common to all living cells. But further investigation has shown interesting differences with regard to the ability of a species to provide other essential trace-constituents. For example, rats and mice make their own ascorbic acid whereas guinea-pigs and man need to have it supplied in the food. Various bacteria differ enormously from one another in their requirements for vitamins, amino-acids and heavy metals. The chemotherapeutic action of the sulphonamide drugs depends on the essentiality of *p*-aminobenzoic acid for many pathogenic bacteria, whereas this acid (the uptake of which is blocked by sulphonamides) is not required by human beings. Some of the trace-substances required by bacteria differ in chemical detail from those required by mammals For example, some types of bacteria utilize a form of vitamin B_{12} which contains bound adenine, whereas humans can use only a B_{12} in which the adenine is replaced by dimethyl-benziminazole. Likewise it has

been demonstrated that many of the enzymes in one-celled forms of life are chemically different from mammalian enzymes performing the same functions (e.g. hexokinases, alcohol oxidases, glucose phosphorylases).

Finally, the quantitative aspect should be considered. Even when two species use similar metabolic pathways, they do not necessarily use them to the same degree. Thus pathogenic trypanosomes metabolize glucose two thousand times faster than man. Hence carbohydrate metabolism is likely to be more vulnerable in these parasites than in their hosts, and this may explain the greater sensitivity of parasites to arsenical drugs.

The Mechanism of Toxicity

Structural Specificity

For the most part, toxic action is concerned with a high degree of chemical specificity: a very small change in the architecture of a molecule is often enough to abolish its biological action. This is true even when the change consists solely of removing a group from one part of a molecule and placing it in another part (i.e. isomerism). For example among the three *p*-aminobenzene-
sulphonamides,
therapeutic. Si
and only one of the seven
bacterial. In each case, the reason for this specificity is known, as we shall see.

It is often asked how anything so small and so inert as a single methyl-group can influence the biological activity of a large molecule. Sometimes the most evident effect is a favourable change in solubility, and it is worth remembering that a lipophilic group like $-CH_3$ can in certain cases actually increase solubility in water. This is seen in the series. sulphadiazine, sulphamerazine, sulphamezathine, which have respectively 0, 1 and 2 methyl-groups in the pyrimidine ring, and which are arranged above in order of increasing aqueous solubility. I believe that this is a crystal-lattice effect

In other cases, methyl-groups change the biological characteristics of a drug by increasing its strength as a base. For example, in the triphenylmethane series, antibacterial action

increases in the following series (arranged in order of increasing alkylation): Doebner's Violet; Malachite Green; Brilliant Green. Apparently only the cations are antibacterial, the percentage ionization at pH 7·3 being 2, 28 and 80 per cent respectively.

Perhaps the commonest reason for drastic changes in biological activity is the fact that adding or subtracting a methyl-group interferes with goodness of fit between the substance and the appropriate receptor-molecules in the cell. For example, the activity of the vitamin thiamine (aneurin) drops to 1/22 of its

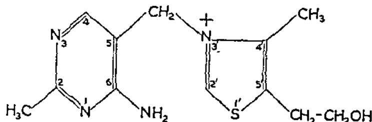


Fig. 1.

former value if a methyl-group is removed from the 2-position (see Fig. 1). If the 4'-methyl-group is removed, the activity drops to 1/82, and if an extra methyl is inserted in the 2'-position, all biological activity is lost. It is known that, in the cell, thiamine is phosphorylated by one enzyme and thereby becomes the coen-

is not in any way unusual. It must often be likewise with drugs which are in many cases (e.g. physostigmine) known to act by blocking enzymes. Fortunately, wider structural variations are permissible in an enzyme inhibitor than in a coenzyme or substrate.

Metabolite Analogues

This class of drugs is easily understood in the light of the foregoing remarks. Douglas, Haldane and Haldane (1912) clearly showed that the poisonous nature of carbon monoxide depends upon its structural resemblance to oxygen (compare the molecules C=O and O=O), so that haemoglobin became saturated

with the former to the exclusion of the latter. Since 1910 it has also been discovered that many enzymes can be inhibited by chemicals which are structurally analogous to their normal substrates. The first application of this knowledge to the chemistry of drugs occurred in 1926 when Stedman showed that the alkaloid physostigmine acted by blocking the enzyme cholinesterase which has the important function of destroying acetylcholine shortly after its liberation from nerve-endings. Thus physostigmine was clearly shown to function by causing the patient to have an overdose of his own acetylcholine. This knowledge led to the simplification of the structure of physostigmine so that only those portions essential for blocking cholinesterase were left. The resemblance of this structure to that of acetylcholine thus became clearer and it was gradually realized that useful drugs could be made by designing molecules to resemble natural cell-constituents. This realization was accelerated by Woods's discovery (Woods, 1940) that sulphanilamide (Fig. 2)—and hence all the sulphonamides—act against bacteria by displacing a structurally similar cell-constituent (*p*-aminobenzoic acid, Fig. 3) from some loosely held combination, e.g. in an enzyme-complex. The structural similarity (Figs. 2 and 3) consists in the fact that a weakly acid and a weakly basic group are in each case attached to a benzene ring and are situated, relative to one another, in the *para* position.

The great practical success of the sulphonamides in therapy is due to the fact that pathogenic bacteria will die if their metabolism of *p*-aminobenzoic acid is interfered with, whereas mammals do not utilize this acid. Attempts to make other 'metabolite analogue' drugs have not often been successful because the science of comparative biochemistry has not yet provided us with sufficient examples of species-differences in metabolism. Nevertheless the successful new antimalarial, 'Daraprim', was evolved (Hitchings, Elion, Van der Werff and Falco, 1948) along these lines as a metabolite analogue for folic acid. Similarly Unna's 'Lethidrone' (Unna, 1943), which is a simple homologue of morphine, is in regular use to rouse patients from overdoses of this alkaloid. Towards the end of the war, a clinically satisfactory antimalarial, phenylpantophenone (closely related

to an essential cell-constituent, pantothenic acid), was evolved along strictly rational lines, but it was not commercially produced because it was not actually superior to known drugs. The drug B.A.L. (British Anti-Lewisite), which is proving so useful in cases of poisoning by arsenic, antimony, mercury and gold, was evolved along somewhat similar lines of reasoning.

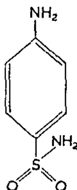


Fig. 2.

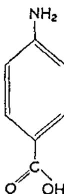


Fig. 3

It seems safe to predict that metabolite analogues will provide us with many useful medicaments eventually. However the difficulties to be overcome are more formidable than was commonly supposed ten years ago. Those interested should read D. W. Woolley's *A Study of Antimetabolites* (Woolley, 1952).

Ionization

The biological action of ions is usually quite different from that of the corresponding molecules. This is not surprising because it is a familiar fact that ions and molecules have quite different chemical reactions. For example, when ascorbic acid is allowed to stand in the air, the neutral molecule oxidizes fastest if iron is present, the mono-anion if copper is present, and the dianion if metals are carefully excluded. Only two factors govern the percentage ionization of a substance, the pK_a and the pH at which the experiment is conducted. Quite small changes in structure can greatly influence the pK_a (negative logarithm of ionization constant). Moreover, quite small changes in pH (e.g. 0·5 unit)

with the former to the exclusion of the latter. Since 1910 it has also been discovered that many enzymes can be inhibited by chemicals which are structurally analogous to their normal substrates. The first application of this knowledge to the chemistry of drugs occurred in 1926 when Stedman showed that the alkaloid physostigmine acted by blocking the enzyme cholinesterase which has the important function of destroying acetylcholine shortly after its liberation from nerve-endings. Thus physostigmine was clearly shown to function by causing the patient to have an overdose of his own acetylcholine. This knowledge led to the simplification of the structure of physostigmine so that only those portions essential for blocking cholinesterase were left. The resemblance of this structure to that of acetylcholine thus became clearer and it was gradually realized that useful drugs could be made by designing molecules to resemble natural cell-constituents. This realization was accelerated by Woods's discovery (Woods, 1940) that sulphanilamide (Fig. 2)—and hence all the sulphonamides—act against bacteria by displacing a structurally similar cell-constituent (*p*-aminobenzoic acid, Fig. 3) from some loosely held combination, e.g. in an enzyme-complex. The structural similarity (Figs. 2 and 3) consists in the fact that a weakly acid and a weakly basic group are in each case attached to a benzene ring and are situated, relative to one another, in the *para* position.

The great practical success of the sulphonamides in therapy is due to the fact that pathogenic bacteria will die if their metabolism of *p*-aminobenzoic acid is interfered with, whereas mammals do not utilize this acid. Attempts to make other 'metabolite analogue' drugs have not often been successful because the science of comparative biochemistry has not yet provided us with sufficient examples of species-differences in metabolism. Nevertheless the successful new antimalarial, 'Daraprim', was evolved (Hitchings, Elion, Van der Werff and Falco, 1948) along these lines as a metabolite analogue for folic acid. Similarly Unna's 'Lethidrone' (Unna, 1943), which is a simple homologue of morphine, is in regular use to rouse patients from overdoses of this alkaloid. Towards the end of the war, a clinically satisfactory antimalarial, phenylpantophenone (closely related

show that neither inorganic iron nor 8-hydroxyquinoline was toxic separately. The reason why none of the six isomerides of Fig. 4 are antibacterial is that none of them have (for structural reasons) the possibility of binding metallic ions. There is good reason to believe that the toxic action of the complex (Fig. 5)

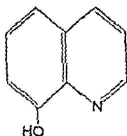


Fig. 4.

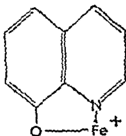


Fig. 5.

takes place inside the cell (where ferrous ions penetrate in only the most carefully regulated amounts) and that it consists of catalysing the oxidative destruction of some group essential in the cell's metabolism (Albert, Gibson and Rubbo, 1953).

Other metal-binding substances which we are now studying include aureomycin and the tuberculostatic drug isoniazide.

The Covalent Bond

As long ago as 1900 Ehrlich noted that: 'If alkaloids, aromatic amines, antipyretics or aniline dyes be introduced into the animal body, it is a very easy matter by means of water, alcohol or acetone to remove all these substances quickly and easily from the tissues.' He concluded that drugs, in general, are able to act on vital cell constituents without forming what we should call covalent bonds, because such bonds are not readily broken at body-temperatures and hence would not release the active principles.

At the present time, we recognize only a few examples where drugs act by forming covalent bonds. Arsenical drugs form one case in point, but the relevant bond (As–S) is abnormally easily broken as covalent bonds go. All useful arsenical drugs are either arsenoxides, as Fig. 6, or are converted by the body to arsenoxides These then act, as Voegtlin showed in 1925, by combining

can change a substance from a substantially non-ionized to a substantially ionized state if the pH and the pK_a are numerically similar, as often happens.

Some classes of drugs (phenols, hypnotics) act best when not ionized, apparently because ionization adversely affects their distribution. Many other substances act best when fully ionized. Some years ago, my colleagues and I discovered that the acridine antibacterials owed their usefulness largely to the cations present. In the course of this work we synthesized over 100 new acridines and showed that those which are poorly ionized at pH 7·3 have little bacteriostatic action whereas those that are highly ionized are invariably highly antibacterial. Incidentally, this work led to the discovery of acridine antibacterials superior to any known at that time. One of these, aminacrine (5-aminoacridine), has now been adopted by the British Pharmacopoeia (Albert, Rubbo, Goldacre, Davey and Stone, 1945). Later, we extended this work to other chemical series and we found that any large flat molecule was antibacterial if substantially ionized (cationically) at the pH of the test. The necessity for largeness (over 30 sq. Å) and flatness are apparently connected with conditions for adsorption on the biologically active group (on the exterior of the cell) which is injured by these cationic drugs (Albert, Rubbo and Burvill, 1949).

Heavy Metals

Of recent years, the importance to the cell of traces of the ions of heavy metals has been increasingly recognized. Thus cobalt, zinc, manganese and copper all play essential parts in bodily metabolism. Possibilities for selective toxicity arise through depriving the uneconomic cell of trace metals or, conversely, increasing the proportion until toxic quantities are reached.

For some years, my colleagues and I have been investigating substances, known as chelating agents, which can combine with traces of metallic ions in such a way that their ordinary inorganic properties are strongly modified For example, 8-hydroxyquinoline (Fig. 4) can combine with traces of ferrous cations in ordinary bacteriological broth to give the complex (Fig. 5), which turns out to be rapidly toxic for bacteria. We were able to

show that neither inorganic iron nor 8-hydroxyquinoline was toxic separately. The reason why none of the six isomerides of Fig. 4 are antibacterial is that none of them have (for structural reasons) the possibility of binding metallic ions. There is good reason to believe that the toxic action of the complex (Fig. 5)

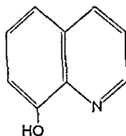


Fig. 4.

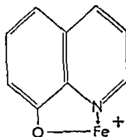


Fig. 5.

takes place inside the cell (where ferrous ions penetrate in only the most carefully regulated amounts) and that it consists of catalysing the oxidative destruction of some group essential in the cell's metabolism (Albert, Gibson and Rubbo, 1953).

Other metal-binding substances which we are now studying include aureomycin and the tuberculostatic drug isoniazide.

The Covalent Bond

As long ago as 1900 Ehrlich noted that: 'If alkaloids, aromatic amines, antipyretics or aniline dyes be introduced into the animal body, it is a very easy matter by means of water, alcohol or acetone to remove all these substances quickly and easily from the tissues.' He concluded that drugs, in general, are able to act on vital cell constituents without forming what we should call covalent bonds, because such bonds are not readily broken at body-temperatures and hence would not release the active principles.

At the present time, we recognize only a few examples where drugs act by forming covalent bonds. Arsenical drugs form one case in point, but the relevant bond (As–S) is abnormally easily broken as covalent bonds go. All useful arsenical drugs are either arsenoxides, as Fig. 6, or are converted by the body to arsenoxides. These then act, as Voegtlin showed in 1925, by combining

can change a substance from a substantially non-ionized to a *substantially ionized state* if the pH and the pK_a are numerically similar, as often happens.

Some classes of drugs (phenols, hypnotics) act best when not ionized, apparently because ionization adversely affects their distribution. Many other substances act best when fully ionized. Some years ago, my colleagues and I discovered that the acridine antibacterials owed their usefulness largely to the *cations* present. In the course of this work we synthesized over 100 new acridines and showed that those which are poorly ionized at pH 7 3 have little bacteriostatic action whereas those that are highly ionized are invariably highly antibacterial. Incidentally, this work led to the discovery of acridine antibacterials superior to any known at that time. One of these, aminacrine (5-aminoacridine), has now been adopted by the British Pharmacopoeia (Albert, Rubbo, Goldacre, Davey and Stone, 1945). Later, we extended this work to other chemical series and we found that any large flat molecule was antibacterial if substantially ionized (cationically) at the pH of the test. The necessity for largeness (over 30 sq. Å) and flatness are apparently connected with conditions for adsorption on the biologically active group (on the exterior of the cell) which is injured by these cationic drugs (Albert, Rubbo and Burvill, 1949).

Heavy Metals

Of recent years, the importance to the cell of traces of the ions of heavy metals has been increasingly recognized. Thus cobalt, zinc, manganese and copper all play essential parts in bodily metabolism. Possibilities for selective toxicity arise through depriving the uneconomic cell of trace metals or, conversely, increasing the proportion until toxic quantities are reached.

For some years, my colleagues and I have been investigating substances, known as chelating agents, which can combine with traces of metallic ions in such a way that their ordinary inorganic properties are strongly modified For example, 8-hydroxyquinoline (Fig. 4) can combine with traces of ferrous cations in ordinary bacteriological broth to give the complex (Fig. 5), which turns out to be rapidly toxic for bacteria. We were able to

When molecular concentrations alone are considered, a 400-fold difference is seen, but this falls to a 2-fold difference if thermodynamic activities are compared (it should perhaps be explained that the word activity is used here in its physical and not biological sense).

TABLE 2. The Narcosis of Tadpoles
(Brink and Posternak, 1948)

Substance	Concentrations for equi-narcotic action mols./litre	Thermodynamic activity of *equi-narcotic solutions*
Alcohol	0 41	0 028
Butanol	0 02	0 021
Octanol	0 0001	0 028
Acetone	0 28	0 030
Ether	0 03	0 040
Chloroform	0 001	0 019

These considerations led Ferguson to suggest the use of the rare gas xenon as a general anaesthetic. In 1952 it was successfully used at the University of Iowa's Hospital in a number of cases. Good relaxation was obtained and the conclusion was reached that this inert gas was a satisfactory anaesthetic.

CONCLUSION

In this subject of selective toxicity much more is known about toxicity than selectivity. However the subject has now reached the stage where it deserves study as an entity. Hence it is hoped that it will secure a firm place in the medical curriculum and receive increased attention as a topic of research in medical schools.

REFERENCES

ALBERT, A. (1951). *Selective Toxicity*. London.
ALBERT, A, GIBSON, M and RUBBO, S. D. (1953). *Brit. J. exp Path* 34, 119.
ALBERT, A., RUBBO, S. D and BURVILL, M. (1949). *Brit J exp. Path* 30, 159.
ALBERT, A., RUBBO, S D, GOLDACRE, R, DAVEY, M. and STONE, J. (1945) *Brit. J. exp Path* 26, 160

reversibly with essential mercapto-groups in the parasite to give substances such as Fig. 7. There are some indications that penicillin acts by forming a covalent bond of the more usual non-reversible character.



Fig. 6.

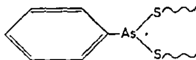


Fig. 7.

Many substances are administered in a masked form, where a covalent bond has to be broken before the true drug is released. The administration of pentavalent arsenicals in order to produce arsenoxides in the c.s.f., where arsenoxides would not ordinarily penetrate, is a case in point. The mild analgesic phenacetin is really *p*-acetamidophenol masked by ethylation of the —OH. *p*-Acetamidophenol is too toxic to be given as such, but is liberated at a suitable rate by the breakdown of phenacetin.

Structural Non-Specificity

So far, all the examples dealt with have involved considerable structural specificity. It is quite otherwise with the hypnotics and general anaesthetics as Overton (1901) and Meyer (1899) pointed out. As no connection exists between structure and biological action for this class of substance, these authors suggested that depressant action on the nervous system is proportional to the distribution of the substances between olive oil and water. This has not proved a good correlation in practice. A genuine advance was made in 1939 when Ferguson pointed out that depressant activity was roughly proportional to the percentage saturation of the substance in the vehicle (air or water) in which it was administered. Thus a substance poorly soluble in water is likely to be more active, molecule for molecule, than one that is more soluble. These conclusions were derived from thermodynamic principles, and increased accuracy is obtained by comparing thermodynamic activities, rather than percentage saturations. How it works out in practice can be gleaned from Table 2.

III

Recent Progress in Antibiotics

SIR ALEXANDER FLEMING

Antibiotics in Clinical Medicine

Research in antibiotics is proceeding in many directions. Following the success of first penicillin and later streptomycin, many teams of workers have been engaged in trying to discover new antibiotics and they have met with considerable success.

The antibiotic boom started with penicillin, and the search for new ones first took place among the moulds, but although many new antibiotics were discovered none of them have been suitable for therapeutic purposes. From the moulds the search extended up and down the plant kingdom from the flowering plants to the bacteria, and in every class antibiotic-producing plants have been found, but the richest source has been the lowest classes—the streptomycetes and bacteria, especially the former.

The work of Waksman and his associates (Schatz, A., Bugie, E. and Waksman, S. A, 1944) in connection with streptomycin attracted attention to streptomycetes, and almost all of the most recent antibiotics have been produced by this group. We have had aureomycin, teramycin and chloromycetin—all wide spectrum antibiotics—which have been extensively used, and we have had neomycin and viomycin which have not come into general use. Then, very recently, we have erythromycin and magnamycin which are still on trial but which seem to have a promising future. All these are elaborated by different streptomycetes as well as many others which are without therapeutic value.

BALDWIN, E. (1948). *Comparative Biochemistry*. Cambridge.
BRINK, F. and POSTERNAK, J. M. (1948). *J. cell. comp. Phys.* **32**, 211.
DOUGLAS, C. G., HALDANE, J. S. and HALDANE, J. B. S. (1912). *J. Physiol.* **44**, 275.
FERGUSON, J. (1939). *Proc. Roy. Soc.* **127B**, 387.
HITCHINGS, G. H., ELION, G., VAN DER WERFF, H. and FALCO, E A. (1948). *J. biol. Chem.* **174**, 765.
MEYER, H. (1899). *Arch exp. Path. Pharmak.* **42**, 109, 119.
OVERTON, E. (1901). *Studien uber die Narkose*. Jena.
STEDMAN, E. (1926). *Biochem J.* **20**, 719
UNNA, K. (1943). *J. Pharmacol* **79**, 27.
VOEGTLIN, C. (1925). *Physiol. Rev.* **5**, 63.
WOODS, D. D. (1940). *Brit. J exp. Path.* **21**, 74.
WOOLLEY, D. W (1952). *A Study of Antimetabolites*. New York.

It is the only antibiotic in common use which has been economically synthesized, and now all the chloromycetin is made in a chemical laboratory and not by fermentation.

These three antibiotics, aureomycin, terramycin and chloromycetin, are known as the wide spectrum antibiotics because they have a very wide range of action on bacteria. They act on the Gram positive and Gram negative bacteria and on the spirochaetes. Their action on the tubercle bacillus is not effective. They also act on the Rickettsia of typhus and other diseases and on some of the larger viruses such as psittacosis and lymphogranuloma inguinale.

They are all readily absorbed when administered orally. They kill off a large proportion of the organisms in the gut. In many diseases they are all effective and it is difficult, from this point of view only, to say in such conditions which is the drug of choice. Chloromycetin passes the blood brain barrier much more readily and appears in the spinal fluid in higher concentration than the others. In typhoid fever and other Salmonella infections it is the only really effective antibiotic. Recently, however, it has been found that after the administration of chloromycetin there has appeared a blood dyscrasia which has been fatal. This has happened especially after long or repeated administration of the drug. While, therefore, chloromycetin is a very effective remedy in many conditions this drawback has to be borne in mind.

During the last few years two other antibiotics derived from streptomycetes have had a considerable clinical trial. These are neomycin and viomycin. They are both active on the tubercle bacillus but with both of them there are drawbacks which outweigh their advantages so that they have not come into general use.

In the last year, two more streptomycete productions have appeared which are at present being tested clinically and which appear to have considerable promise. These are erythromycin and magnamycin.

Erythromycin is a base which forms salts with organic and inorganic acids. It is stable at the degrees of pH found in the tissues. Its spectrum includes especially the Gram positive organisms but also some Gram negative ones and some viruses.

The search still goes on but each discovery of an antibiotic makes it more difficult to find a new one which has any advantage over those already in use.

Penicillin has a relatively limited range of action. Its spectrum includes most of the Gram positive organisms, the Neisseria, the spirochaetes and a few others, but the wide range of Gram negative organisms are for the most part unaffected. The great merit of penicillin is its extraordinary power on the organisms that are sensitive, and especially its absence of toxicity—it is almost impossible to give an overdose. Its drawback is its instability in solution, but that matters little for systemic administration as it is freely soluble and solutions are readily made at the time of injection.

Streptomycin acts on both Gram positive and Gram negative organisms. It is effective in many cases of urinary infection with coliform organisms but it cannot deal with systemic infections like typhoid fever or other Salmonella infections. Its special value has lain in its action on the tubercle bacillus and there is no need to stress the difference it has made in many tuberculous conditions. Just as penicillin gave hope of life to sufferers from subacute bacterial endocarditis, so streptomycin gave a similar hope to those with tuberculous meningitis.

Streptomycin has to be injected—when taken orally very little is absorbed.

It has two drawbacks as compared with penicillin. Unless care was exercised in the dosage there were toxic symptoms, especially a specific toxicity to the eighth nerve. The other drawback is the very rapid development of resistance in bacteria which are not destroyed.

Aureomycin and terramycin are somewhat similar in chemical constitution, but there is an atom of chlorine in the aureomycin molecule which is absent in terramycin. They are very similar in their antibacterial activity, although there are considerable differences in their properties One of these differences is that aureomycin is very unstable in solution.

Chloromycetin (chloramphenicol) was originally a product of a culture of a streptomyces (*S. venezuela*). It has a relatively small molecule containing one chlorine atom and a nitrobenzene ring.

Polymyxin is derived from *B. polymyxa*. It has also been described under the name Aerosporin. Its spectrum is completely different from the two preceding bacterial antibiotics. It acts on Gram negative bacilli and the Gram positive cocci are insensitive. It has something of the same nephrotoxicity as bacitracin so in systemic use great care has to be taken regarding dosage and the urine must be watched for signs of disturbance of kidney function. It has been used a good deal as a local application for burns or other extensive surface lesions especially with a view to ridding these of *Proteus* or *B. pyocyaneus*.

These are the antibiotics which are used clinically, but the search for new antibiotics goes on vigorously. For example, the September 1952 number of *Antibiotics and Chemotherapy* contained descriptions of four new antibiotics all derived from streptomyces.

An enormous amount of work has to be done before a new antibiotic is issued to the general medical public. Six separate stages can be listed.

1. A biologist has to discover that one micro-organism produces something which antagonizes other microbes. He has to develop culture media in which this substance is freely produced and he has to investigate the antibiotic 'spectrum', i.e. ascertain which organisms are sensitive and which are insensitive.

2. A chemist has to purify the active substance as far as he can. If he succeeds in purifying it, he tries to discover its formula and if possible synthesize it.

3. A pharmacologist has to study the action of the purified product. He investigates its absorption and excretion. Especially he has to study its toxicity, and not only its acute toxicity after a single dose but also the possibility of organic damage after a long continued course of smaller doses. (Many antibiotics which at first seemed non-toxic were found to cause serious damage to the kidneys or other organs after continued administration) Then he sees the effect of the drug on experimental infections in animals.

4. If the antibiotic passes the tests of the pharmacologist, the physician sees whether it can cure disease in man. He has from the biologist a list of the sensitive organisms. He has from the pharmacologist the dose which is effective in animals and also

It is especially recommended in infections with staphylococci, streptococci, and pneumococci. *Haemophilus influenzae* and *H. pertussis* are also sensitive, and it is suggested that it may be of value in whooping cough and diphtheria. It is given orally in doses of 0·8 to 2·0 grammes daily in divided doses.

Side effects are uncommon but large doses may give rise to nausea, vomiting and diarrhoea.

Magnamycin has a spectrum somewhat similar to penicillin but *in vitro* it affects enterococci which tend to be resistant to penicillin and the Rickettsia which are resistant. The pyogenic cocci are inhibited by a concentration of about 0·2 μ/gramme per c.c., C. diphtheria by 0·04 μ/gramme and gonococcus and haemophilus influenzae by 0·02 μ/gramme. Toxicity to animals is very slight and the experimental therapeutic effects have been excellent. It can be given orally or by intramuscular or intravenous injection.

It will be very interesting to see the results of further clinical trials with these two drugs. As there seems to be no cross resistance with penicillin, they may be of the greatest value in the treatment of infections by penicillin-resistant staphylococci which are so common in hospitals today.

Three antibiotics derived from bacteria are available. These are tyrothricin, bacitracin and polymyxin.

Tyrothricin is the only antibiotic in use which was described before the penicillin boom. Dubos described it in 1939. It acts only on Gram positive bacteria and it is so toxic that it is useless for systemic treatment. It is available for local treatment.

Bacitracin is elaborated by a strain of *B. licheniformis*. It has a spectrum somewhat similar to penicillin and there is no cross resistance between the two. With some batches especially there has been a certain amount of toxicity to the kidney and for this reason it has not come into very general use. Many patients have, however, been treated with intramuscular injections of 40,000 to 80,000 units daily in divided doses with comparatively little toxic reaction. One interesting point about bacitracin is that there is a synergism between it and penicillin enabling results to be obtained with much smaller doses of each than would be necessary if they were used alone.

Originally the stock dosage was 15,000 units every 3 hours and good results were obtained, although very often penicillin could not be detected in the blood after 2½ hours. Later, as supplies increased, larger doses were given at longer intervals and even if there was a gap of some hours when penicillin had apparently disappeared from the blood between doses the results were good.

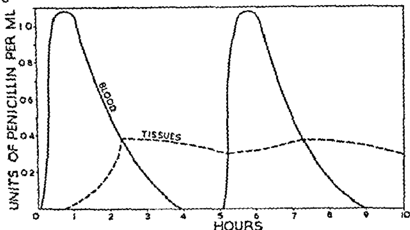


FIG. 1.

It is impracticable when a patient is not in hospital to give more than two injections a day and patients and doctors alike prefer only one. Many patients have been treated with doses of 100,000 units twice a day and frequently the treatment has been successful. After such a dose penicillin can be found in the blood for 5 or 6 hours, so there is a gap of some 6 hours before the next injection during which time there is less than a therapeutic level of the drug in the blood.

We have to remember, however, that the microbes which are infecting the body are seldom in the blood stream. They are in the tissues, and it is worth a moment's consideration to see what sort of concentration of penicillin there would be in the tissues. Fig. 1 illustrates what could happen.

When an *intramuscular injection is given, there is a very* rapid rise in the penicillin content of the blood. Then the penicillin diffuses from the blood to the tissues. The blood level

information on absorption and excretion so that he knows whether he has to give it orally or parenterally. He has to observe the rate of excretion and the level of the drug in the blood so that he may know how frequently to repeat the dose, and he has to adjust the dose so that a therapeutic level is maintained in the blood for a sufficient time.

5. If, from the observations of the physician, the drug appears valuable, the pharmaceutical manufacturer has to devise methods of production on a large scale. With penicillin this was difficult as so many new problems were encountered, but now problems connected with the growth of the organism in large containers are better understood and progress is much more rapid. It was four or five years after penicillin was first purified before satisfactory large-scale manufacturing methods were devised, but only some nine months elapsed between the discovery of terramycin and its issue to the public.

6. Someone gives the antibiotic a name. Penicillin was so named because it was elaborated by a penicillium and in general the name of the genus or species of the organism producing the antibiotic is indicated in the name allotted to it.

Is it necessary in Antibiotic Treatment that the Blood should constantly contain a Therapeutic Level of the Drug?

In this connection we had best deal with penicillin for it is concerning penicillin that we have the most information available.

TABLE I

Dose in units	Time (in hours) during which penicillin can be detected in the blood after intramuscular injection
15,000	2½–3
30,000	3
50,000	4
100,000	5–6
250,000	7–8
500,000	9–10
1,000,000	12

With different doses of sodium penicillin the drug can be found in the blood for times approximately as shown in Table I.

Originally the stock dosage was 15,000 units every 3 hours and good results were obtained, although very often penicillin could not be detected in the blood after 2½ hours. Later, as supplies increased, larger doses were given at longer intervals and even if there was a gap of some hours when penicillin had apparently disappeared from the blood between doses the results were good.

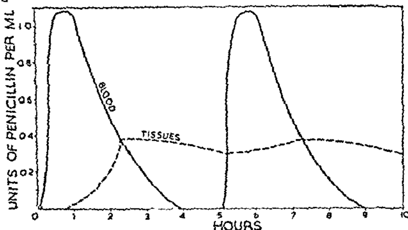


Fig. 1.

It is impracticable when a patient is not in hospital to give more than two injections a day and patients and doctors alike prefer only one. Many patients have been treated with doses of 100,000 units twice a day and frequently the treatment has been successful. After such a dose penicillin can be found in the blood for 5 or 6 hours, so there is a gap of some 6 hours before the next injection during which time there is less than a therapeutic level of the drug in the blood.

We have to remember, however, that the microbes which are infecting the body are seldom in the blood stream. They are in the tissues, and it is worth a moment's consideration to see what sort of concentration of penicillin there would be in the tissues. Fig 1 illustrates what could happen.

When an intramuscular injection is given, there is a very rapid rise in the penicillin content of the blood. Then the penicillin diffuses from the blood to the tissues. The blood level

rapidly falls, and in an hour or two there is equilibrium between the blood and tissues. The blood level continues to fall and then there is diffusion back from the tissues to the blood. There must always be some period during which penicillin exists in the tissues in effective concentration after it ceases to be detectable in the blood.

Infected tissues are usually oedematous and in such tissues it will take longer for the drug to diffuse in to its maximum, but, equally, it will take longer to diffuse out.

We had a patient with a streptococcal infection and a very oedematous leg. Following penicillin by intramuscular injections (in the buttock) the penicillin levels were tested at intervals in the blood and in tissue fluid from the oedematous leg. The blood levels fluctuated as usual, rising soon after the injection and falling to nothing before the next injection. The tissue fluid levels were almost constant, only varying between 0·4 and 0·5 unit per ml. during the whole day.

Florey and Heatley (1945) have shown that penicillin persisted in the fluid exuded from septic wounds for a considerable time after it had disappeared from the blood.

More recently, Ungar (1950), working with rabbits, obtained similar findings. He created a necrotic focus by the injection of turpentine into the muscles of the hind leg. Then penicillin was injected intramuscularly into the muscles of the other leg. Blood and exudation from the necrotic area were examined at intervals with the following results.

TABLE 2

Dose 30,000 units intramuscular	Time (in hours) during which penicillin detectable in	
	Blood	Exudate
Aqueous penicillin	3	8–9
Oil wax ,,	12	18
Procaine ,,	24	31

We see then that penicillin persists in the tissues (especially inflamed tissue) for some time after it has disappeared from the blood.

We must not therefore insist that penicillin shall be in the blood all the time. We know that the clinical results have very often been excellent when there is no measurable amount of penicillin in the blood for some time between injections and the observations cited give one explanation of this.

The interval, however, between the injections must not be too long, and it is reasonable to suppose that with a quickly growing organism the interval should be shorter than with a slow growing organism. I would suggest that in dealing with the common septic organisms it would not be wise to give less than 500,000 units if two doses a day are given, or less than 1,000,000 units if only one dose of soluble penicillin is given in 24 hours.

Chemical Developments in Penicillin

Penicillin is an acid and readily makes salts with bases—inorganic or organic. Originally only simple salts like sodium, potassium or calcium were used. These were very soluble and when injected in solution they were rapidly absorbed and rapidly excreted. For this reason the injections had to be repeated frequently unless very large doses were given. Many methods were devised to delay absorption and so prolong the action. The first successful method was to incorporate the penicillin in oil and beeswax. There were, however, technical difficulties and the oil wax mixture gave way to organic salts.

Early in penicillin treatment when the drug contained many impurities there was often considerable pain at the site of injection and frequently procaine was added to the penicillin to prevent this pain. It was sometimes noticed that in this mixture a precipitate appeared but this precipitation did not destroy the efficacy of penicillin. Further investigation showed that procaine formed with penicillin a compound which was only very slightly soluble in water. When a suspension of this compound was injected intramuscularly it very slowly dissociated so that after a single injection a therapeutic concentration of penicillin could be found in the blood for 24 hours. This made it possible to do effective penicillin therapy with only one daily injection. It was further found that when to a suspension of procaine penicillin in oil a certain amount of aluminium monostearate was added,

rapidly falls, and in an hour or two there is equilibrium between the blood and tissues. The blood level continues to fall and then there is diffusion back from the tissues to the blood. There must always be some period during which penicillin exists in the tissues in effective concentration after it ceases to be detectable in the blood.

We had a patient with a streptococcal infection and a very oedematous leg. Following penicillin by intramuscular injections (in the buttock) the penicillin levels were tested at intervals in the blood and in tissue fluid from the oedematous leg. The blood levels fluctuated as usual, rising soon after the injection and falling to nothing before the next injection. The tissue fluid levels were almost constant, only varying between 0·4 and 0·5 unit per ml. during the whole day.

Florey and Heatley (1945) have shown that penicillin persisted in the fluid exuded from septic wounds for a considerable time after it had disappeared from the blood.

More recently, Ungar (1950), working with rabbits, obtained similar findings. He created a necrotic focus by the injection of turpentine into the muscles of the hind leg. Then penicillin was injected intramuscularly into the muscles of the other leg. Blood and exudation from the necrotic area were examined at intervals with the following results.

TABLE 2

Dose 30,000 units intramuscular	Time (in hours) during which penicillin detectable in Blood	Exudate
Aqueous penicillin	3	8–9
Oil wax ,,	12	18
Procaine ,,	24	31

We see then that penicillin persists in the tissues (especially inflamed tissue) for some time after it has disappeared from the blood.

Turner and Drew, 1945) showed that penicillin-resistant staphylococci could be developed in this way and that these retained their resistance to penicillin through many generations. Later it was shown that this could happen with many organisms. With penicillin the development of resistance was slow and proceeded in steps, the organisms becoming more resistant as they were grown in increasing concentrations of the chemical.

This slow development of resistance characteristic of penicillin is seen also in most of the other useful antibiotics, aureomycin, terramycin, and chloromycetin, but it is otherwise with streptomycin. Here a high degree of resistance may be developed in a very short time.

The appearance of penicillin-resistant forms has so far only been of importance in relation to one common organism—the staphylococcus. In the 'thirties, before penicillin was used clinically, I found a few staphylococcal strains which appeared to be highly resistant. They could not have developed by the medicinal use of penicillin.

After penicillin had come into general use in hospitals, Barber (Barber and Rozwadoski-Dowzenko, 1948) published figures showing that of the staphylococci isolated in one hospital some 50 per cent were resistant to penicillin. Later figures show an even higher incidence of resistant organisms.

Against this we have several sets of figures showing that among people not in hospital the incidence of penicillin-resistant staphylococcus is about 10 per cent. There is then a great difference in the incidence of penicillin-resistant staphylococci in hospitals and out of hospitals. Why is this? The simple answer is that it is a hospital infection. We know from history, even recent history, how easy it is to spread infection in hospital. In hospital many patients are treated with penicillin. That kills off the sensitive staphylococci but not the few insensitive ones. These insensitive ones find they have the field to themselves so they grow and multiply. They get into the air and the dust of the wards and infect not only the other patients but the nurses and attendants, who can spread the infection. It is the old story of hospital infection which is difficult to prevent completely but which, with care, can be cut down to small limits.

penicillin could be detected in the blood for three days or more.

Naturally there was never a high level in the blood. There could not be for there was only slow dissociation of the penicillin from the compound and the free antibiotic was rapidly excreted by the kidneys. However, if it was desired to have an initial high level in the blood it was easy to mix a dose of soluble sodium penicillin with the procaine salt and many manufacturers have issued such a mixture.

More recently other organic compounds have appeared. Jensen and his associates (Jensen, Dragsted and Klair, 1950) in Denmark introduced the hydroiodide of the diethylaminoethanol ester which like procaine penicillin forms a depot from which the drug is only slowly absorbed. The interesting feature of this compound is that in animals with an inflammatory condition of the lungs a much higher concentration of penicillin is

In the last few months there has been introduced another organic compound of penicillin dibenzylethylenediamine dipenicillin G. This is marketed under various trade names and after a single injection of 600,000 units a measurable amount of penicillin appears in the blood for some 14 days. The titre is never high—at the most in the region of 0·1 unit per ml. but after 14 days there may still be 0·02 or 0·03 unit per ml. which is enough to deal with very sensitive organisms.

Then again there is allylmercaptomethyl penicillin, sold as Penicillin O, which is stated to give much less reaction in penicillin-sensitive patients.

Doubtless other compounds will appear, each having some advantage in a special direction.

Development of Resistance of Bacteria to Antibiotics

It was well known that bacteria grown in sublethal concentrations of a chemical could develop a resistance to the chemical. Soon after penicillin was introduced clinically, Todd (Todd,

Turner and Drew, 1945) showed that penicillin-resistant staphylococci could be developed in this way and that these retained their resistance to penicillin through many generations. Later it was shown that this could happen with many organisms. With penicillin the development of resistance was slow and proceeded in steps, the organisms becoming more resistant as they were grown in increasing concentrations of the chemical.

This slow development of resistance characteristic of penicillin is seen also in most of the other useful antibiotics, aureomycin, terramycin, and chloromycetin, but it is otherwise with streptomycin. Here a high degree of resistance may be developed in a very short time.

The appearance of penicillin-resistant forms has so far only been of importance in relation to one common organism—the staphylococcus. In the 'thirties, before penicillin was used clinically, I found a few staphylococcal strains which appeared to be highly resistant. They could not have developed by the medicinal use of penicillin.

After penicillin had come into general use in hospitals, Barber (Barber and Rozwadoski-Dowzenko, 1948) published figures showing that of the staphylococci isolated in one hospital some 50 per cent were resistant to penicillin. Later figures show an even higher incidence of resistant organisms.

Against this we have several sets of figures showing that among people not in hospital the incidence of penicillin-resistant staphylococcus is about 10 per cent. There is then a great difference in the incidence of penicillin-resistant staphylococci in hospitals and out of hospitals. Why is this? The simple answer is that it is a hospital infection. We know from history, even recent history, how easy it is to spread infection in hospital. In hospital many patients are treated with penicillin. That kills off the sensitive staphylococci but not the few insensitive ones These insensitive ones find they have the field to themselves so they grow and multiply. They get into the air and the dust of the wards and infect not only the other patients but the nurses and attendants, who can spread the infection. It is the old story of hospital infection which is difficult to prevent completely but which, with care, can be cut down to small limits.

There is evidence that the simultaneous administration of two antibiotics may prevent the appearance of resistant strains of bacteria and in practice this has been demonstrated clearly by the combination of streptomycin and P.A.S. in the treatment of tuberculosis. Much work remains to be done before we know the best combination of antibiotics for this and other purposes.

It is often assumed that bacteria which become resistant to an antibiotic remain resistant permanently, but very interesting work has been done in Great Britain and in America showing that if some resistant organisms are exposed to the action of various chemicals they may regain their sensitivity to one or more antibiotics. This change, however, does not seem to keep pace with the spread of the resistant forms.

Evaluation of Antibiotics in the Treatment of Disease in Man

Every one of the antibiotics in common use has had such dramatic results in some conditions that there has been no need for accurate statistical analysis to show them to be valuable.

There were conditions, however, where strictly controlled trials were necessary and in this connection British medicine and the Medical Research Council have played a major part.

Let us take, for instance, subacute bacterial endocarditis. In the pre-penicillin era it was almost invariably fatal and it was shown that penicillin treatment could cure it. The question of how best to use penicillin in this condition remained. It needed a controlled trial and here the Medical Research Council stepped into the breach. They organized such a trial comparing the results obtained by various systems of dosage in a number of different centres. Large doses over a short period failed. Smaller doses over a longer period were often successful but the greatest number of successful results were obtained by comparatively large doses over quite a long period. A final result of this trial has appeared in a very interesting article by Cates and Christie (1951).

The M.R.C. trials of streptomycin and other antitubercular drugs have likewise been of the greatest value in the evaluation of these drugs in the treatment of various forms of tuberculosis.

Similar trials are in progress in regard to the use of antibiotics

in adult pneumonia, whooping cough and gastro-enteritis of infants.

Such organized and strictly controlled trials are essential, first, to ascertain whether the drug has an appreciable effect on the particular infection and, secondly, to determine the best method of using the drug or drugs. Many authors publish results of observations in which there are no controls and these results can be most misleading.

The Combination of Different Antibiotics

When penicillin was introduced it was the only antibiotic which could be used systemically but even when there was no competing antibiotic there were attempts to combine it in treatment with the sulphonamides and Bigger (1946) showed that the addition of sulphathiazole to penicillin enhanced its bactericidal

with combinations of antibiotics has become very important.

There is experimental work showing that when bacteria are exposed to a number of antibacterial agents, resistant strains do not readily develop. There is considerable clinical evidence that in tuberculosis it is normal for resistant strains to appear in about 4 weeks of streptomycin treatment but if the streptomycin is combined with para-amino-salicylic acid the development of these resistant strains is prevented or retarded.

When two antibiotics are combined three other things can happen.

1. Each may act independently so that the addition of one to another is merely addition.

2. The antibiotics combined may have an antibacterial action far in excess of a mere additive effect. One or both synergizes the other.

3. *One antibiotic may inhibit the action of another.*

Jawetz and Gunnison (1952) have divided the antibiotics into two groups:

Group I—penicillin, streptomycin, bacitracin

Group II—aureomycin, terramycin, chloromycetin.

From experiments *in vitro* and *in vivo* they have formulated the following rules.

TABLE 3

Mixtures	Results
Group I with Group I	Never antagonism—possibly synergism
Group II with Group II	Never antagonism or synergism Occasionally additive
Group I with Group II	If organism sensitive to Group I often antagonism If organism resistant to Group I often synergism

With the combination of group I and group II the result depends to some extent on the microbe. There is one very illustrative experiment quoted. An organism very sensitive to bacitracin was exposed to a mixture of bacitracin and terramycin. The result was worse than with bacitracin alone. The same organism was, by repeated subculture in sublethal doses of bacitracin, rendered relatively resistant to the drug. When this semi-resistant organism was treated with the bacitracin-terramycin mixture the result was much better than with either drug alone. Thus the same combination of antibiotics showed both antagonism and synergism on the same organism according to the state of the culture.

If we accept this finding as a working hypothesis we see that the problem of using combinations of antibiotics is far from being a simple one. It seems to be safe to combine members of group I which are never antagonistic to each other, so long as care is taken to avoid the possible toxic effects of streptomycin and bacitracin. There is little advantage in combining members of group II for at best they are merely additive

The serious danger may come when a member of group I is combined with a member of group II. There may be antagonism, so that the result is worse than if only a single antibiotic is used. This is especially likely if the organism is reasonably sensitive to the group I drug.

In some countries mixtures of antibiotics are manufactured and sold. I hope this practice does not become general. It is so

easy to combine the drugs for any particular patient at the moment of administration.

Unskilled practitioners have a great love for a blunderbuss mixture and if such were available in the antibiotic field they would use it for everything—often with worse results than they would get by the use of a single drug.

Toxic Action of Antibiotics

Toxic effects of three types may be observed.

1. The drug may have a specific toxic effect on some tissues, for instance streptomycin and dihydrostreptomycin have a direct toxic action on the eighth nerve if given in large doses over more than a very short period.

Chloromycetin has recently been shown to have in some patients a very serious toxic action on the blood-forming tissues leading to a fatal aplastic anaemia.

Terramycin and aureomycin taken orally may cause nausea and vomiting or diarrhoea.

Bacitracin and polymyxin, unless given in very small doses, have a direct toxic action on the kidney. This is the most serious objection to their general use.

2. The drug may upset the bacterial equilibrium of some part of the body. Normally in our mouths, throats and intestinal tract we have a very considerable bacterial flora which adjusts itself so that we accept it as the normal condition. If a large proportion of these bacteria are destroyed by an antibiotic their place is taken by other organisms resistant to that antibiotic and these may cause disease. Penicillin lozenges kill off most of the streptococci in the mouth and sometimes this is followed by 'black tongue' or thrush.

The wide spectrum antibiotics kill off most of the bacteria of the gut and in many cases there is a great overgrowth of fungi, especially monilia—sometimes with serious results.

3. The patient may be allergic to the drug or may by exposure to the drug acquire a sensitivity.

Soon after penicillin was introduced as a therapeutic measure it was found that some people were allergic to the drug. Usually the manifestations were slight, such as urticaria, but occasionally

there were more serious symptoms. Some cases have been reported where persistence with the treatment in the face of such manifestations has had a disastrous result.

I heard of one during my recent visit to Greece. A man with chronic bronchitis treated himself with penicillin (200,000 units). After the first dose there was slight and transient dyspnoea. Three more doses were taken at approximately weekly intervals and after each there was increasing dyspnoea. He then consulted a doctor who warned him not to take any more penicillin but he disobeyed the advice and had another dose which was rapidly followed by a fatal dyspnoeic attack.

Apart from the primary sensitivity, a sensitivity can be acquired by repeated exposure to one of these drugs. Many nurses and others who had day after day to deal with streptomycin developed a sensitivity of the skin resulting in a dermatitis

The recently reported toxic effects of chloromycetin on the blood-forming organs are probably due to a sensitization. They have occurred especially after very long continued repeated administration.

It is probable that with all the antibiotics there are individuals who develop such a sensitivity especially after repeated local applications of the drug. This is one of the main drawbacks to the local application of these drugs. It is also the chief reason why it may be important not to use them for quite trivial complaints. In countries where they are freely sold to the public they have been used indiscriminately and I have given above an instance of a fatal result from self-medication with penicillin which is normally a drug with extremely little toxicity.

Conclusion

We now have a number of effective antibiotics and in them we have powerful weapons against most of the common infections, and we shall have new antibiotics. There remain many problems which are yet unsolved.

Antibiotics are the products of laboratory research and the physicians should seek the co-operation of the laboratory. In some cases the laboratory can be of material aid in determining

whether the drugs are being absorbed in sufficient quantity by measuring the blood level but it is especially useful in the selection of the most suitable antibiotic for use in any particular infection.

In antibiotic therapy the association of the clinician and the laboratory worker is a very happy one.

REFERENCES

BARBER, M. and ROZWADOSKI-DOWZENKO, M. (1948). *Lancet*, ii, 641.

BIGGER, J. W. (1946). *Lancet*, i, 81.

CATES, J. E. and CHRISTIE, R. V. (1951). *Quart. J. Med.* **20**, 93.

DUBOS, R. J. (1939). *J. Exp. Med.* **70**, 1 and 11.

FLOREY, M. E. and HEATLEY, N. G. (1945). *Lancet*, i, 748.

JAWETZ, E. and GUNNISON, J. B. (1952). *Antibiotics Chemotherapy*, **2**, 243.

JENSEN, K. A., DRAGSTED, P. S. and KLAER, I. (1950) *Ugeskr. Laeg* **112**, 1075

SCHATZ, A., BUGIE, E. and WAKSMAN, S. A. (1944). *Proc Soc. exp. Biol*, N.Y. **55**, 66.

TODD, E. W., TURNER, G. S. and DREW, L. G. W. (1945). *Brit. med. J.* **1**, 111.

UNGAR, J. (1950). *Lancet*, i, 56.

IV

Virus Adaptability in Relation to Human Disease

WILSON SMITH

ADAPTABILITY to environment is one of the most important characteristics of living organisms. Any organism totally lacking such capacity would be doomed to early extinction. More than temporary adaptations, however, are required to permit the survival of a species as distinct from the survival of a mere individual. The species must evolve a mechanism for the transmission of changed characters through a number of successive generations. This implies a capacity for constant evolutionary development.

In this lecture I do not propose to discuss the various mechanisms by which heritable changes are brought about in the first place and subsequently handed on from generation to generation. The generally accepted view is that any permanent change of character depends upon the selection of a mutant which arises spontaneously by pure chance and which happens to have an increased survival value over the parent form. Whether that is the only pattern of evolutionary development is immaterial to my present purpose What I wish to discuss is the effect which virus mutability has upon the human diseases caused by various viruses and its important bearing on some of the practical problems which confront us in the field of human medicine. There can be no doubt that the phenomenon of virus mutation has very great practical significance, for its scientific investigation has already yielded, and is likely to yield in future, handsome dividends in the form of methods of control of some of the virus infections.

Virus Lability

Whilst it is probably true to say that all living organisms are subject to changes based upon genetic mutations they differ widely amongst themselves in the frequency with which they exhibit such changes. The viruses in particular display an extraordinary range of variability in this respect. Some of them are relatively stable probably because they have already achieved such a satisfactory mechanism for ensuring species survival that any further mutation conferring an increased survival value is extremely unlikely to occur. Mumps virus provides an example of this relative stability with the result that the human disease, mumps, has persisted unchanged through the centuries. At the other end of the scale the influenza viruses are so extremely labile that aberrant strains may be produced almost at will in the laboratory, whilst the natural disease presents a clinical and epidemiological picture of bewildering complexity. It is paradoxical that whilst virus lability may be the chief reason for man's failure to conquer a disease in one case it may, in another case, be the essential prerequisite for the development of an effective means of control.

Range of Virus Adaptations

With a very labile virus there is scarcely any type of activity which may not be altered by genetic modification. Indeed, one of the difficulties of the research worker is to detect such changes in his laboratory strains in order to guard against the risk that they may vitiate the applicability of his experimental results to problems in clinical medicine and epidemiology. Burnet (1945) drew attention to this risk in the case of influenza when he wrote: 'One has an uneasy feeling that perhaps all our current influenza work is being done with something which has ceased to be the pathogenic virus with which we ought to be primarily concerned', and he emphasized the point by adding, 'Little progress would have been made in dealing with pneumococcal pneumonia if only rough pneumococci had been available for study'. This does not in the least detract from the significance and importance of laboratory investigations, for it is only by studying the changes which occur in the course of laboratory

IV

Virus Adaptability in Relation to Human Disease

WILSON SMITH

ADAPTABILITY to environment is one of the most important characteristics of living organisms. Any organism totally lacking such capacity would be doomed to early extinction. More than temporary adaptations, however, are required to permit the survival of a species as distinct from the survival of a mere individual. The species must evolve a mechanism for the transmission of changed characters through a number of successive generations. This implies a capacity for constant evolutionary development.

In this lecture I do not propose to discuss the various mechanisms by which heritable changes are brought about in the first place and subsequently handed on from generation to generation. The generally accepted view is that any permanent change of character depends upon the selection of a mutant which arises spontaneously by pure chance and which happens to have an increased survival value over the parent form. Whether that is the only pattern of evolutionary development is immaterial to my present purpose. What I wish to discuss is the effect which virus mutability has upon the human diseases caused by various viruses and its important bearing on some of the practical problems which confront us in the field of human medicine. There can be no doubt that the phenomenon of virus mutation has very great practical significance, for its scientific investigation has already yielded, and is likely to yield in future, handsome dividends in the form of methods of control of some of the virus infections.

that it demonstrated the possibility of non-antigenic modifications affecting the biological activities of a virus, and thus lent support to the idea that sudden increases of virulence, acquisition of new tissue affinities and possibly extensions of host range of a virus may depend upon architectural alterations of a similar type.

However, whatever may be their physical or chemical basis, there is abundant evidence that both antigenic and non-antigenic virus mutations occur in nature just as they do in the laboratory. They occur with variable frequency depending chiefly upon the basic stability or lability of the virus in question. Many of them are likely to be of such a nature that they are without effect upon the fundamental reactions between host and parasite, but those which affect virulence, tissue affinities or transmission potentialities are bound to modify the host-parasite *relationship in one way or another. Furthermore any antigenic* modification, even when not associated with a change of biological activity, is likely to be important from the point of view of prevention and control of the disease. Indeed the association or dissociation between antigenic mutations and functional changes of a virus may be the crucial factor in determining the type of control measure likely to be effective.

In order to illustrate the manifold effects of this plasticity I propose to consider four specific virus diseases, smallpox, yellow fever, influenza and poliomyelitis, with special reference to certain major aspects of infection. namely, the clinical and epidemiological features of the disease, methods of disease transmission and preventive measures.

Smallpox

In smallpox we have an example of a virus which is extremely stable antigenically but which, fortunately for us, is liable to mutations affecting its virulence for man and its experimental host range. During the last few centuries the disease has exhibited striking fluctuations of severity and distribution. Of course, one must not lose sight of the multiplicity of factors, other than virus mutation, which can affect the clinical and epidemiological features of a disease. Early accounts of epidemics

experiments that one can hope to unravel the complexities of virus behaviour in nature. The wide range of the modifications of influenza virus which have been encountered during the course of laboratory investigations is shown in Table 1.

TABLE 1. Variable Characters of Influenza Virus

Variable character	Behaviour affected by virus modification
Antigenic constitution	Epidemiological patterns.
	Efficacy of vaccines.
Infectivity	Host range.
	Adaptability to laboratory animals.
	Tissue affinities.
Enzymic activity	Susceptibility to serum inhibitors.
	Speed of elution from R.B.C.s (and ? tissue cells).
Haemagglutination activity	G. pig/Fowl R B.C. ratio.
	Susceptibility to serum inhibitors.
Toxicity	? Severity of clinical symptoms.
Morphology	
Resistance	Persistence in nature.

A point of very great importance is that mutations may affect either the virus constitution or its functional activities. In the last analysis any change of behaviour must be dependent upon some alteration of constitution but in many cases this is quite undetectable and may involve merely a rearrangement of structure rather than any gain or loss of actual virus components. Yet a change of this sort may have a much greater effect upon the interactions between virus and human host than a readily detectable antigenic change. Recently in my own laboratory a variant of the classical PR8 strain was isolated during routine egg passage which differed from the parent strain in haemagglutination behaviour without any detectable change in infec- [illegible], and we [illegible] merely [illegible] e haem- agglutinin receptors on the virus surface (Smith *et al.*, 1951). This particular modification, if it occurred in nature, would be unlikely to have any significance as far as the human disease is concerned, but I regard the occurrence as of some importance in

with the result that smallpox would cease to be a particularly severe disease. A factor militating against such a transformation is the widespread use of prophylactic vaccination which raises a barrier against that free spread of infection which is necessary for such natural selection to occur.

The close similarity between the basic pathology of smallpox and that of several of the animal pock diseases, coupled with the antigenic relationships of the viruses concerned, makes it practically certain that the group of pock viruses evolved from a common ancestry. We know nothing definite about the order of their evolution but there is undoubtedly a particularly close relationship between the viruses of smallpox and cowpox and there is at least a reasonable possibility that one may have developed as a mutant of the other by adaptation to the new host species. It is significant that cowpox is readily transmissible to man on contact with infected animals, and there are some authorities who hold that cowpox in the cow may result from contact with human cases of smallpox or vaccinia. If, indeed, one virus represents a mutation from the other the evolution may have proceeded equally well in either direction though the much more violent host reaction in the case of smallpox suggests that smallpox virus represents the more recent development.

Fortunately, epidemic smallpox is entirely dependent upon man-to-man transmission; there are no insect or animal vectors and no animal reservoirs of the virulent man-adapted virus. But that is not to say that further developments along such lines are impossible The experimental adaptation of smallpox virus to lower animals is not difficult though it seems to be invariably associated with a decline of virulence for man. As we shall see later in connection with other virus diseases this decline of virulence for the primary host is by no means an inevitable consequence of adaptation to an alternative host species, and it is an interesting intellectual exercise to work out the events which might follow were smallpox virus to extend its natural host range without undergoing any concomitant change in respect of its infectivity, communicability and virulence for man. Quite certainly the whole epidemiological pattern of smallpox would be changed and urgent new problems of effective control would

in Europe provide a picture of a fulminating disease with high mortality and extraordinarily high infectivity. The shift to a disease of somewhat milder character is generally attributed to improvement of social conditions, the development of some degree of herd immunity and the employment of prophylactic vaccination. Apart however from the periodic fluctuations dependent upon such factors there has arisen a stable type of mild smallpox, alastrim, which results from infection with a virus variant. Virus strains isolated from alastrim cases show no antigenic differences from strains isolated from cases of classical virulent smallpox. But there is a consensus of opinion that the two disease types, smallpox major and minor, are caused by viruses which represent two distinct and stable entities. The mortality rates in the great epidemics recorded in history appear to have been at least 30 per cent. In more recent times the reported cases in India during the period 1926–30 showed a mortality rate of 42 per cent. But during roughly the same period the mortality rate for cases in the U.S.A. was as low as 0 9 per cent This cannot be attributed to any difference of racial susceptibility, because some of the outbreaks in the U.S.A. were of the smallpox major type with high mortality. The contrast is undoubtedly due to the dominance of the alastrim virus in America during the period in question. That mortality rate is governed chiefly by the type of infecting virus is also indicated by the fact that, by and large, outbreaks breed true to type; this is to say that nearly all cases in an alastrim epidemic are relatively mild and lack some of the clinical features characteristic of variola major

Here then is a clear illustration of the effect which virus adaptation may have upon the clinical and epidemiological features of a disease In this case the two disease types are still extant and may occur side by side in any community but it is unlikely that under perfectly natural conditions they would remain so indefinitely. The mild alastrim type virus has almost certainly an increased survival value over the virulent smallpox major type at least in highly civilized countries with high standards of public health, so that in the course of time it would probably succeed in becoming completely dominant,

There is a good deal of evidence that variations of virulence occur constantly during the various transmission cycles which the virus undergoes in nature, and that natural selection has led to the establishment of some of the variants as relatively stable strains. With few exceptions African strains are more pathogenic for rhesus monkeys than are South American strains, whilst for the marmoset it is the American strains which are much more highly pathogenic. Different strains also show extremely variable potentialities of infecting *Aedes aegypti* mosquitoes. There is no correlation between virulence for man and for other species of host but there can be no doubt that comparable virus changes occur which do determine the degree of infectivity and virulence for human beings. Infection may result in clinical disease of any grade of severity ranging from transient malaise to fulminant disease with the full range of classical symptoms. Completely inapparent infections are also common. We know of course that in yellow fever host factors including genetic constitution play a very important part in deciding the nature of the response to infection, but they may on occasion be quite overshadowed by the intrinsic virulence of the parasite. Serological tests in the Sudan in 1935 showed regions in which about 80 per cent of the inhabitants had been previously infected without any history of the disease, but in the same districts outbreaks of classical severe yellow fever occurred in 1940–41. Similarly in Brazil, which has suffered some of the most devastating epidemics on record, outbreaks of a mild febrile illness, originally attributed to influenza, were subsequently proved to be caused by yellow fever virus of abnormally low virulence.

The epidemiology of yellow fever is largely determined by the mode of transmission of the disease. The tissue tropisms of the virus are such that free excretion of virus from the diseased host does not occur; this excludes the possibility of direct transfer of infection from case to case. Virus transfer depends upon participation by a mosquito vector which functions not merely as a mechanical carrier but as a true alternate host of the virus. A further point of interest is that the only period during the course of the illness in which the mosquito can be infected by feeding on the human subject is the brief period of viraemia. Also the

arise. However, though such an adaptation is theoretically possible, it is not very likely to occur because the inherent qualities of the pock viruses appear to favour mutations along different lines.

If our supposition that cowpox and smallpox evolved either one from the other or both from a common ancestry is correct, they present an excellent example of the way in which naturally occurring adaptations may provide a very efficient means of disease control. It only required the penetrating observation of Jenner to open the way for the employment of one virus as a prophylactic against infection with the other. The success of this early use of a naturally occurring phenomenon soon had further consequences. It was found that the changes which could be induced in smallpox virus by artificial adaptations to experimental animals tended to follow a constant pattern which resulted in virus variants having the same desirable qualities as the natural cowpox virus. Although the original Jennerian vaccinia virus may have been derived from cowpox virus many of the strains now used throughout the world for human vaccination have undoubtedly been derived from virulent smallpox virus by a series of artificial adaptations. The really remarkable feature of these numerous conversions of smallpox virus to vaccinia virus is the consistency of the association between reduction of infectivity for man, increase of infectivity for calf and retention of initial antigenic constitution. The first character justifies deliberate infection of human beings with a living disease agent, the second facilitates large-scale vaccine production, and the third ensures almost complete cross-protection.

Yellow Fever

The virus of yellow fever resembles smallpox virus in one very important characteristic, namely its remarkable antigenic stability. All strains of the virus which have been so far isolated, whether from man, monkey or mosquito, are immunologically identical. The adaptative changes to which it is subject in nature or during the course of laboratory manipulations affect chiefly its virulence and its tissue tropisms. The effects of this adaptability upon the major aspects of infection which we selected for discussion present many new and interesting features.

viscerotropism. The mouse neurotropic virus has been used, and indeed is still being used, on a large scale by the French for human immunization, but English and American authorities consider it is too dangerous, an opinion supported by the fact that there have been a number of fatal cases of infection due to the vaccination. The problem was therefore to reduce viscerotropism without increasing neurotropism. The virus strain labelled 17D was finally evolved which fulfills this requirement satisfactorily and which has been employed for the immunization of millions of individuals without any untoward event. The parent virus was a highly virulent pantropic strain isolated from a human case in rhesus monkeys. It was first modified by tissue culture in mouse embryo medium. This resulted in great reduction of viscerotropism but had no effect on neurotropism. The culture medium was changed by substitution of chick embryo for mouse embryo. Later a further modification of the medium, in which chick embryos from which brains and spinal cords had been removed were employed, resulted in a culture virus strain with the qualities necessary for safe and satisfactory human vaccination. *Both viscerotropism and neurotropism were reduced* but infectivity and antigenic constitution remained. The present-day vaccine consists simply of homogenized chick embryos which have been inoculated with a seed virus of the 17D strain. The evolution of the two vaccines in current use is outlined in Fig. 1.

It is interesting that primary isolation of pantropic yellow fever virus from the animal host has never been accomplished by chick embryo tissue culture. This indicates that in the case of this virus the satisfactory reduction of both viscerotropism and neurotropism for the production of a suitable immunizing strain involves a series of stepwise mutations rather than a single mutation. A further attribute is also requisite before a living yellow fever virus can be adopted for human immunization with complete confidence, namely reduced transmissibility. Otherwise every vaccination would involve the risk of the vaccinated individual initiating an epidemic. Indeed the result might well be a major outbreak of high virulence for it has been shown that serial monkey passage of an attenuated strain results in restoration

mosquito remains non-infective for about twelve days after feeding on infected blood. A man-mosquito cycle with such limitations would, in itself, probably be inadequate to ensure survival of the virus in nature, but a large reservoir of infection is provided by jungle monkeys. These facts make it extremely unlikely that man was the original host. However, no matter which species had the honour of serving in that capacity, it is certain that a remarkable series of virus adaptations or mutations must have occurred before the present host-parasite relationships were firmly established. In gaining the power of infectivity for alternate host species the virus must have lost the capacity of direct transmission within the primary host species. Is it possible that further adaptations may occur in future which will tend to alter the epidemiological pattern of the disease? My own answer would be that not only is it possible but that sooner or later it is inevitable. I can see no valid reason why the acquisition of, say, pneumotropism with resultant spread of yellow fever by droplet infection should be regarded as impossible.

Unlike smallpox virus, yellow fever virus has no close relative, analogous to the virus of cowpox, which could be used for immunization. Although naturally occurring strains exhibit widely different degrees of virulence so that some of them apparently cause very mild infections of man, there are none which could be accepted as safe for such a purpose. Nevertheless yellow fever immunization with living virus vaccine is now practised on a large scale and represents one of the great triumphs of virus research and preventive medicine. This practical control of yellow fever is entirely dependent upon the fact that adaptative virulence changes can be impressed upon the virus without altering its antigenic constitution. All strains of the virus possess two major tissue affinities which are conveniently labelled viscerotropism and neurotropism, and these can be modified by appropriate experimental manipulations. Theiler (1930) first showed that serial intracerebral passage of the virus in mice resulted in progressive increase of neurotropism, coupled with progressive decrease of viscerotropism. Findlay and Clarke (1935) demonstrated that the change is reversible, for intrahepatic inoculation of rhesus monkeys led to a restoration of

viscerotropism. The mouse neurotropic virus has been used, and indeed is still being used, on a large scale by the French for human immunization, but English and American authorities consider it is too dangerous, an opinion supported by the fact that there have been a number of fatal cases of infection due to the vaccination. The problem was therefore to reduce viscerotropism without increasing neurotropism. The virus strain labelled 17D was finally evolved which fulfills this requirement satisfactorily and which has been employed for the immunization of millions of individuals without any untoward event. The parent virus was a highly virulent pantropic strain isolated from a human case in rhesus monkeys. It was first modified by tissue culture in mouse embryo medium. This resulted in great reduction of viscerotropism but had no effect on neurotropism The culture medium was changed by substitution of chick embryo for mouse embryo. Later a further modification of the medium, in which chick embryos from which brains and spinal cords had been removed were employed, resulted in a culture virus strain with the qualities necessary for safe and satisfactory human vaccination. Both viscerotropism and neurotropism were reduced but infectivity and antigenic constitution remained. The present-day vaccine consists simply of homogenized chick embryos which have been inoculated with a seed virus of the 17D strain. The evolution of the two vaccines in current use is outlined in Fig 1.

It is interesting that primary isolation of pantropic yellow fever virus from the animal host has never been accomplished by chick embryo tissue culture. This indicates that in the case of this virus the satisfactory reduction of both viscerotropism and neurotropism for the production of a suitable immunizing strain involves a series of stepwise mutations rather than a single mutation. A further attribute is also requisite before a living yellow fever virus can be adopted for human immunization with complete confidence, namely reduced transmissibility. Otherwise every vaccination would involve the risk of the vaccinated individual initiating an epidemic. Indeed the result might well be a major outbreak of high virulence for it has been shown that serial monkey passage of an attenuated strain results in restoration

mosquito remains non-infective for about twelve days after feeding on infected blood. A man-mosquito cycle with such limitations would, in itself, probably be inadequate to ensure survival of the virus in nature, but a large reservoir of infection is provided by jungle monkeys. These facts make it extremely unlikely that man was the original host. However, no matter which species had the honour of serving in that capacity, it is certain that a remarkable series of virus adaptations or mutations must have occurred before the present host-parasite relationships were firmly established. In gaining the power of infectivity for alternate host species the virus must have lost the capacity of direct transmission within the primary host species. Is it possible that further adaptations may occur in future which will tend to alter the epidemiological pattern of the disease? My own answer would be that not only is it possible but that sooner or later it is inevitable. I can see no valid reason why the acquisition of, say, pneumotropism with resultant spread of yellow fever by droplet infection should be regarded as impossible.

Unlike smallpox virus, yellow fever virus has no close relative, analogous to the virus of cowpox, which could be used for immunization. Although naturally occurring strains exhibit widely different degrees of virulence so that some of them apparently cause very mild infections of man, there are none which could be accepted as safe for such a purpose. Nevertheless yellow fever immunization with living virus vaccine is now practised on a large scale and represents one of the great triumphs of virus research and preventive medicine. This practical control of yellow fever is entirely dependent upon the fact that adaptative virulence changes can be impressed upon the virus without altering its antigenic constitution. All strains of the virus possess two major tissue affinities which are conveniently labelled viscerotropism and neurotropism, and these can be modified by appropriate experimental manipulations. Theiler (1930) first showed that serial intracerebral passage of the virus in mice resulted in progressive increase of neurotropism, coupled with progressive decrease of viscerotropism. Findlay and Clarke (1935) demonstrated that the change is reversible, for intrahepatic inoculation of rhesus monkeys led to a restoration of

extremely improbable that a suitable strain would ever evolve naturally as in the case of smallpox to await exploitation by an inspired observer. Reduced transmissibility by the mosquito vector would almost certainly confer reduced survival value on any first-step mutant which would therefore tend to die out before further mutational steps could take place.

There is however a danger in emphasizing the distinction between immunizing mutants that have arisen and become stabilized in nature, and those that have been developed by laboratory manipulations. The difference is more apparent than real and it is indeed largely artificial. We have no means of controlling the direction of genetic mutation. What we can do, and what was done for the production of 17D, is to provide the conditions for the selection and maintenance of those naturally occurring mutants which happen to suit our purpose. With other viruses like those of the pock group suitable conditions happen to have been provided by nature. This is not to say that we shall always remain thus limited. We do know already of several means by which the frequency of mutation can be increased; by X-radiation, treatment of the virus with nitrogen mustard and so on. Sooner or later, type of mutation will also be controllable. When that happens the effect upon disease control may be incalculable.

Influenza

The two viruses we have considered so far, namely those of smallpox and yellow fever, are good examples of the dependence of specific prophylaxis upon a variant which can only arise, whether in nature or in the course of laboratory work, because of the intrinsic lability of the parent viruses. The influenza viruses display much greater intrinsic lability, yet influenza stands in striking contrast to both smallpox and yellow fever in that its lability is the major reason for failure to obtain a satisfactory immunizing agent. I think that the reason for this is that mutations of the influenza viruses nearly always involve, or are associated with, changes of antigenic constitution whereas, as we have already seen, smallpox virus and yellow fever virus exhibit great antigenic stability. Before proceeding to this aspect

of full virulence and there is no reason to suppose that serial passage through human beings would fail to have the same effect. Fortunately 17D does have reduced transmissibility though it is not certain that this is due to a genetic change of the virus. Mosquitoes fed on vaccinated men, and then after the appropriate time interval fed on susceptible monkeys, fail to

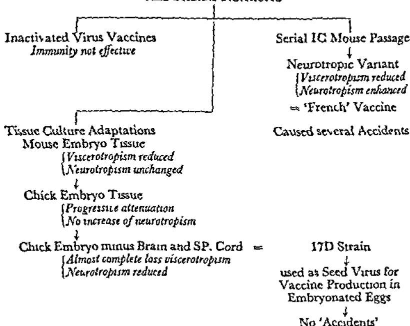


Fig. 1. Yellow Fever Vaccines.

transmit the infection. Similarly mosquito transmission of infection from vaccinated to normal monkeys is not possible. The immersion of mosquito larvae in virus vaccine results in infection of the adult mosquitoes but, even so, the latter are non-infective for monkeys.

These facts indicate that without the modifications of the virus induced by prolonged and purposive laboratory manipulations it is unlikely that successful artificial immunization against yellow fever would ever have become available, for it is

we have come to recognize the remarkable plasticity of the influenza viruses and to discard our earlier conceptions of the fixity of antigenic types.

In any single outbreak the viruses isolated usually conform to a single antigenic strain. Occasionally more than one strain, antigenically distinguishable, are recovered, but the periodic dominance of one strain throughout all affected districts or even countries is the rule. On the other hand the viruses isolated in different epidemics are never identical and the longer the interval between the respective epidemics the greater tends to be the strain divergence. The two major virus types, A and B, are so distinct antigenically that one cannot have arisen as a direct mutant from the other, although both may well have evolved from a common ancestor. But within each of these major types all degrees of antigenic relationship are met with. In the early days of influenza virus research it was thought that there were a limited number of antigenic components, different combinations of which produced the different virus strains. It rather looks now as though there is no limit to influenza virus antigenic variability, for as one strain evolves out of another new antigens seem to evolve to replace older ones. There is now only very slight relationship between current A type strains and the original W.S. virus first isolated in 1933. So different in fact are the strains isolated since 1946 that they have to be classified in a sub-group called A prime. Whether the evolutionary changes ever result in the re-emergence of earlier types we do not know. The possibility is not without importance in connection with the periodic appearance of pandemic influenza.

It is unfortunate that means of virus isolation were not available during the last great pandemic, because had laboratory comparisons of the 1918 virus with the strains isolated from ordinary epidemics been possible, much might have been learnt about the major factors concerned with virulence and communicability. Strictly speaking of course we can never be absolutely certain that the 1918–19 pandemic was caused by a virus related to the known influenza viruses but the circumstantial evidence in support of this view is I think overwhelming. It is thus extremely probable that relatively minor changes in influenza virus

of the question, however, I wish to discuss very briefly the clinical manifestations and the epidemiology of the disease.

In many cases of clinical influenza it is impossible to isolate any virus or to obtain any serological evidence of viral aetiology. It is, therefore, possibl[illegible] some of the variability in symptomatology and epidem[illegible] diversity of aetiological agents. [illegible] majority of ou[illegible] can now be shown, [illegible] another, to be [illegible] symptomatology of one outbreak may [illegible] times the majority of cases are extre[illegible] from a day or two of indisposition; sometimes they are so severe that several days in bed are required. In one epidemic practically all the c[illegible] a long period of convalescence during which nervous [illegible] depression are apparent; in the next [illegible] out of ten cases may enjoy rapid comple[illegible] times the emphasis may be on localized symptoms referable [illegible] involvement of the upper respiratory tract, such as catarrh and sore throat, or upon general symptoms such as malaise, anorexia, limb and back pains and so on. Occasionally an epidemic occurs in which a high proportion of the cases suffer some lung involvement with a correspondingly high mortality rate.

The communicability of the infection also fluctuates from outbreak to outbreak. Many outbreaks remain localized but others spread from end to end of a country, from one country to another or even from continent to continent.

Although much of the variability in the symptomatological and epidemiological pattern of influenza must be attributed to the ever-changing immune status of populations at risk there can be no doubt that virus mutability is also a major factor. The virus strains isolated from different outbreaks are found to behave very differently from each other in laboratory experiments, some of them having peculiarities of initial virulence, toxicity and adaptability to alternate host species which serve as useful marker characteristics. How far such infectivity potentialities are affected by minor changes of antigenic constitution is not known but it is chiefly from the results of antigenic analyses that

addition the epidemic came along only a few weeks after the vaccinations had been completed. A few years later in 1947 other large-scale trials were a complete failure because, in the meantime, the virus had undergone a series of adaptations which resulted in the epidemic strain being of the A prime type for which the vaccine did not contain the effective antigens (Salk *et al.*, 1945; Rasmussen *et al.*, 1947; Francis *et al.*, 1947).

Opinions vary as to the feasibility of overcoming these difficulties, but I have said enough to show how the whole problem of specific immunization is linked up with the idiosyncrasy of virus mutability and how the techniques evolved for the control of one disease may be quite inapplicable to another.

POLIOMYELITIS

Poliomyelitis, the last of the four specific virus diseases which I have chosen to illustrate the effects of virus adaptability, calls attention to certain aspects of the subject which are not covered by the other three infections we have so far considered.

Until recently poliomyelitis virus was considered to be exclusively a pathogen of man, without phylogenetic relationships with any other pathogens, either of man or of lower animals. It was also considered to be obligatory neurotropic and to be one of the most stable viruses known. The discovery of different virus types and especially the adaptations of the Lansing virus to rodents, with the subsequent demonstration of biological relationships between it and other rodent pathogens, demanded modification of these earlier conceptions. It is now customary to speak of the Poliomyelitis group of viruses. Also the weight of evidence now supports the view that manifestations of neurotropism during the course of natural human infections are relatively rare and may represent more or less accidental occurrences. It is certainly established beyond doubt that the individuals showing clinical signs of involvement of the central nervous system represent a mere fraction of the total infected with the virus. Indeed many, perhaps the majority of, infected persons show no sign of disease whatsoever and some of these may excrete the virus in their faeces in large quantity. The factors which determine the epidemic occurrence of paralytic

structure have from time to time in the past produced killing strains of such extraordinary communicability that they have swept over most of the earth's habitable zones and further it is extremely probable that the same thing will happen again in future.

I have already indicated that the antigenic mutability of the influenza viruses is of particular importance in connection with specific prophylaxis. Let me say at once that, quite apart from this fact, virus vaccination against influenza could never be as effective as it is against smallpox and yellow fever. The duration of immunity following natural infection is quite short whereas in the other diseases mentioned it is usually lifelong. In spite of this limitation, however, and even to some extent because of it, a reliable and effective vaccine would prove of enormous benefit to mankind. The death rate from the pandemic of the first world war was reckoned in millions whilst the suffering and economic loss from recurrent epidemics of low mortality are incalculable. It is therefore not surprising that research on vaccination against influenza has been and is still being vigorously pursued.

Experience both with experimental animals and trials in human beings shows quite clearly that it is possible to confer protection by this means. The trouble is, first, to know what virus we need to protect against, and, second, to have available the right vaccine at the right time. The closer the antigenic relationship between vaccinating and challenge strains of virus the more successful is the result of the vaccination. The virus types A and B give no cross-protection whatsoever—the A and A prime types give relatively little. Even within the single A type there are great differences in the degrees of cross-protection exhibited by different strains. It is therefore not surprising that the great majority of field trials have been either failures or inconclusive. The most successful large-scale trial on record is that carried out by the Influenza Commission in the U.S.A. in 1943. In four out of five centres, incidence of influenza in the vaccinated was only about a quarter of that in the unvaccinated controls. The virus responsible for the epidemic turned out to be almost identical with a strain which had been isolated a short time before and which had been incorporated in the vaccine. In

extended backwards to the period of the first world war without significant fluctuation. In 1947 however, there was a really startling rise of incidence. Such a peak in one single year might possibly be caused by various extraneous circumstances but similar high incidence occurred in 1949, 1950 and 1951. This strongly suggests that either the endemic strain was replaced by a mutant in 1947 or that a new virus strain was introduced from abroad. It is indeed difficult to account for such a change on any

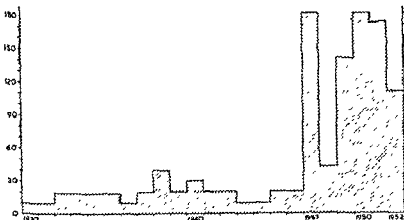


Fig. 2. Poliomyelitis notifications per million of population in England and Wales for the period 1930–52.

other basis. It is extremely unlikely that a change in host susceptibility could have occurred with such dramatic suddenness, and it is perhaps relevant that no comparable phenomenon has occurred in respect of measles or whooping cough during the last thirty years.

The shift of age incidence might equally well be due to an enhanced virulence of the dominant strain, enabling it to overcome the tougher physiological barriers of older children, but the more favoured explanation is that improved social conditions and standards of personal hygiene now protect many infants who in earlier times would have been exposed to infection. Thus children nowadays tend to grow up without the advantage of an acquired immunity. These alternative hypotheses however are by no means mutually exclusive. As Burnet

poliomyelitis from time to time are unknown and are probably multiple, but Burnet considers that the major factor is virus mutation to an unduly invasive type. There is, however, an alternative possible explanation which is not dependent upon mutability of the aetiological agent. A group of viruses, known as the Cocksackie viruses, has been brought to light in recent years. These are intestinal parasites of man which may cause infections bearing a close resemblance to non-paralytic poliomyelitis. It is known that poliomyelitis and cocksackie viruses may occur side by side in one and the same individual, and it has been suggested that paralytic poliomyelitis may represent a combined infection in which the two viruses act together synergistically (Melnick *et al.*, 1951). Although there is no direct evidence that this occurs, the hypothesis derives some support from the discovery, made at Hampstead, that mouse hepatitis, a disease of mice similar to infective hepatitis of man, is caused by two filterable agents acting in combination (Gledhill *et al.*, 1952). This of course opens up the possibility that other diseases of unknown aetiology may eventually be found to depend upon similar virus synergism.

However that may be, it is certainly hard to resist the conclusion that virus mutation must have played a part in the striking change of epidemiological pattern of poliomyelitis which has occurred over the last half-century. Previous to about sixty-five years ago sporadic cases and small outbreaks occurred, but the large and severe outbreaks which nowadays cause us such alarm were unknown. This change was not gradual but appeared suddenly in the Stockholm epidemic of 1887, from then onwards periodic outbreaks of increasing size and severity have occurred. The change is best explicable on the assumption that a strain, or strains, of virus, with increased virulence and more especially increased neurotropism, evolved and became the dominant type.

It now seems as though a *further* mutation in the direction of increased communicability has occurred in the last five years at least as far as England is concerned. Fig 2 shows a graph of poliomyelitis annual incidence in England and Wales during the period 1930 to 1952 Up to 1947 there was relatively little annual fluctuation. Actually the graph might have been

virus evolution, it is obvious that the trend of mutation has been *directly away from strains which might be suitable for artificial* immunization of man and in this respect poliomyelitis stands in striking contrast to smallpox. Natural immunization with strains of low virulence is undoubtedly occurring constantly, but so far we have no means of isolating, characterizing and propagating such strains for the production of suitable living virus vaccines. Also the lack of any means of virus propagation in tissues other than those of the central nervous system of monkeys or mice, has hitherto prevented the purposive selection and modification of suitable mutants by laboratory manipulations in ways similar to those which were so successful in the case of yellow fever. Recently a very important advance was *made by the cultivation* of poliomyelitis virus in human embryo tissue cultures, and very recently this strain has been adapted to monkey testis tissue cul*tures (Enders et al., 1949; Scherer et al., 1951). This may indeed* represent the essential first steps in the required chain of virus adaptations. Intensive efforts are being made to adapt the virus to growth in the chick embryo and should this be achieved a new chapter in poliomyelitis research would open. However, even if this were successfully accomplished it would not necessarily mean that a satisfactory vaccine could be evolved. It would remain to find out whether the adaptations also involved a change of *antigenic constitution which might vitiate the object* in view. We do not yet know how wide is the range of antigenic variation amongst the naturally occurring strains which infect man. The occurrence of second attacks of poliomyelitis in a single individual and the results of cross-immunity experiments in monkeys both indicate that the immunity resulting from infection is not as effective against a heterologous virus strain as against the homologous strain. However the rarity of second attacks *and the low incidence of paralytic polio compared with* the very high incidence of occult infections suggest that strain variation is unlikely to be in any way comparable with that found amongst the influenza viruses. These facts, I suggest, give reasonable grounds for hope that an effective poliomyelitis vaccine will eventually be evolved.

has pointed out, it is by a change of host environment that mutants, better able to cope with the new situation, are selected out to become the new dominant virus type. Thus whilst improvement of social and hygienic conditions may for a time favour the host at the expense of the parasite so that the control of a given disease appears to be in sight, there is a real risk that the devil cast out merely makes room for a worse devil to enter. This is what appears to have happened in the case of poliomyelitis in recent years.

This changed pattern of the disease during the present century may well have resulted from a virus mutation of a character rather more fundamental than mere enhancement of virulence. Burnet (1945) suggests that it may be based upon the acquirement by the virus of an alternative mode of transmission. The progenitor virus was probably a pathogen of rodents or even a harmless intestinal parasite of rodents which was freely excreted in the faeces. Contamination of human food must have been of very frequent occurrence under the conditions of life of our primitive ancestors when hygiene as we know it today was completely unknown. There were therefore ample opportunities for adaptation of the virus to the human host and sooner or later a human strain would develop from a mutant which happened to be favoured by the environment provided by the human intestine. So long as the virus retained low potentiality for invasion of the central nervous system the disease would remain of little import, especially as social habits would ensure the early infection of the majority of infants. Under such circumstances further mutants of a more dangerous type would have little chance of establishing themselves, and paralytic poliomyelitis would remain a somewhat rare and accidental catastrophe. When, however, by rising standards of hygiene, the free spread by faecal contamination was prevented with the consequent loss of early immunization, the stage was set for the selection of any mutant capable of adopting an alternative mode of transmission. In some such way the more modern neurotropic strains capable of spread by either faecal contamination or direct droplet infection probably arose.

If this represents anything like the true picture of poliomyelitis

GLEDHILL, A. W., DICK, G. W. A. and ANDREWES, C. H. (1952). *Lancet*, ii, 509.
MELNICK, J. L., KAPLAN, A. S., ZABIN, E., CONTRERAS, G. and LARKUM, N. W. (1951). *J. exp. Med.* **94**, 471.
RASMUSSEN, A. F., STOKES, J. C. and SMADEL, J. E. (1947). *Amer. J. Hyg.* **47**, 142, 1942.
SALK, J. E., MENKE, W. J. and FRANCIS, T. (1945). *Amer. J. Hyg.* **42**, 57.
SCHERER, W. F., BUTORAC, G. and SYVERTON, J. T. (1951). *Fed. Proc.* **10**, 417.
SMITH, W., WESTWOOD, M. A., WESTWOOD, J. C. N. and BELYAVIN, G. (1951). *Brit. J. exp. Path.* **32**, 422.
THEILER, M. (1930). *Ann. trop. Med. Parasit.* **24**, 249.

New Virus Diseases

Finally I would like to add a brief word about the origin of new virus diseases. The way in which viruses first evolved is quite unknown and remains a fertile subject for speculation and controversy but, given any virus which has succeeded in establishing itself securely as an animal pathogen, there is no difficulty in appreciating how it may undergo further adaptative changes to produce a new human disease. Virus adaptations from species to species must have been going on from time immemorial. There are indications that, at the present time, this process is continuing by the adaptation to man of some of the insect-borne neurotropic viruses of lower animals such as equine encephalomyelitis. So long as insect transmission remains the only mode of spread the human disease is likely to remain sporadic and of relatively minor importance, but the acquirement of alternative mechanisms of virus transmission or the adaptation to man of respiratory infections of animals may be fraught with much graver consequences. It seems to me that psittacosis and the rickettsial disease, Q fever, represent cases of adaptation to man in which the host-parasite relationship is very delicately poised. In either of these cases the evolution at the right moment of a mutant with enhanced virulence and especially enhanced communicability might well result in the establishment of the infection as a major epidemic disease of man. At the moment there is little we can do about it but a scientific appreciation of the possibilities is surely the first step towards control of our own destiny.

REFERENCES

Burnet, F. M. (1945). *Virus as Organism*. Harvard Univ. Press, Cambridge, Mass.

Enders, J. F., Weller, T. H. and Robins, F. C. (1949). *Science*, 109, 85.

Findlay, G. M. and Clarke, L. P. (1935). *Trans. R. Soc. trop. Med Hyg.* 28, 579.

Francis, T., Salk, J. E and Quilligan, J J (1947). *Amer. J. publ. Hlth.* 37, 1013.

In the second half of the last century, the advent of the aniline dye industry, initiated by Perkin's discovery of the aniline process in 1855, saw the first cases of industrial cancer of the bladder in the aniline workers in Germany. Here some relatively pure substances were employed, and the substance *beta*-naphthylamine was thought on clinical grounds to be responsible for some of these bladder tumours. It was not until very much later that this observation was confirmed experimentally; probably because of the variable susceptibility of different species of animals to any particular carcinogen. This is still one of our major difficulties in assessing the possible carcinogenicity of any particular substance for the human subject. Schar in 1930 described bladder tumours in rabbits subjected to inhalation of *beta*-naphthylamine, but the histological evidence was debatable. In 1938 Hueper, Wiley and Wolfe obtained papilloma or carcinoma of the bladder in 13 out of 16 female dogs subjected first to injection and later to oral administration of up to 300 mg. of *beta*-naphthylamine *per diem* during 20 months of experiment.

After the discovery that tar could cause cancer in the skin of mice, Passey in 1922 showed that a benzene extract of soot was similarly active, and Leitch showed, also in 1922, that heavy mineral oils, used in mule-spinning in Lancashire, could also cause tumours in the skin of mice. Thus there was complete correlation between the experimental evidence for carcinogenicity for tar, soot and heavy mineral oils and the clinical observation of industrial cancer.

Very careful studies were made by Murray (1921) and other workers of the Imperial Cancer Research Fund of the various stages in the development of skin cancer in mice following painting with tar. Every stage from chronic inflammatory lesions to frank malignancy was seen and described, and not unnaturally the idea developed that chronic irritation was in some way carcinogenic. However, tar is an extremely complex mixture of substances, many of which are irritants and some of which are carcinogens. It is probable that some degree of inflammatory disturbance is necessary before a carcinogen can act on the previously resting cells, but it does not follow that chronic irritation of a non-specific kind will be followed by cancer.

V

Carcinogenesis

P. R. PEACOCK

As with the birth or beginning of many other things, we know in a practical way how to set about generating cancer; but when we come to the fundamental understanding of the processes involved we know comparatively little. The experimental pathologist can so set the stage that he can be reasonably certain of producing a tumour in an animal of a chosen species, or even in a particular organ or tissue. Yet this feat in itself tells us nothing about the mechanisms involved.

Since the recognition by Percival Pott (1775) of chimney-sweep's cancer of the scrotum as an industrial disease, a number of other hazardous occupations have been correctly associated with certain types of tumour incidence in the workers concerned, and experimental pathologists during the same period corroborated the deductions of their clinical colleagues and have devised a number of experimental hazards for animals comparable in their way with the industrial risks run by some human beings.

Carcinogenic Properties of Tar

Hanau, who in 1889 first successfully transplanted a tumour from one rat to another, also had the idea that tar was likely to be a carcinogen and painted the scrotal skin of rats with this substance but without achieving positive results. Had he chosen mice instead of rats, he might have advanced our knowledge of carcinogenesis by some twenty-five years. It remained for Yamagiwa and Ichikawa in 1915 to make the observation that tar could induce tumours in the ears of rabbits, and two years later Tsutsui showed that the same thing was true for the skin of mice.

particularly when they are endowed, as many of these carcinogens are, with the physical property of fluorescence, one would be able to follow quite [illegible] s in the conversion of a [illegible] as not proved to be the [illegible] way this should be. If we [illegible] ovum we can see nothing to give us an indication of how this one cell can contain within itself the blueprint for a new individual. Similarly, if we examine reacting connective tissue or epithelial cells in [illegible] t with a carcinogen, [illegible] w they are about to [illegible] the presence of the [illegible] place.

[illegible] we are not entirely without information about the preliminary changes in the cell that is destined to become malignant. Pullinger (1940) has shown that certain early changes occur in the epithelial cells of the skin of mice exposed to some of the more potent chemical carcinogens. While these changes are not absolutely pathognomonic, they indicate that some change has taken place in the cells quite early after exposure to the carcinogen—within four or five days—and other work suggests that these changes may be irreversible. For example, it has been shown by Berenblum (1941) that the cells of the skin can be conditioned by substances not themselves carcinogenic, in such a way as to be either more or less prone to develop cancer than normal cells.

In the course of experiments designed to test the effects of chronic irritants, he showed that mustard gas—a very powerful irritant—when applied in extremely dilute solution to the skin of mice, could increase its resistance to the action of chemical carcinogens such as 3:4-benzpyrene. On the other hand, application of croton oil—another potent irritant—to the skin of mice after a single application of benzpyrene, not in itself sufficient to cause tumours with certainty, could [illegible] rs to develop.

[illegible] tio [illegible] hood of a tar [illegible] the appearance of tumours in the traumatized tissue. Rous and his colleagues (Rous and Kidd,

The problem was simplified to some extent by Kennaway's (1924) synthesis of carcinogenic tars from isoprene and acetylene heated at 720°–900° C. in atmospheres of hydrogen. The tars in this instance could only contain hydrocarbons and carbon, since no other elements but hydrogen and carbon were present in the reagents used. This narrowed the field of search to the hydrocarbons which still included a formidable list of possible aromatic and aliphatic substances. Anthracene, one of the aromatic hydrocarbons present in tar, was not found to be carcinogenic; but 1:2:5:6-dibenzanthracene was shown by Kennaway and Hieger (1930) to be a potent carcinogen and to have the same characteristic fluorescence that Mayneord (1927) has associated with the synthetic tars and that Hieger (1930) also found characteristic of carcinogenic heavy mineral oils. 1:2:5:6-dibenzanthracene had been synthesized by Clar in Czechoslovakia in 1929 for reasons unconnected with cancer research. His accurate recording of the physical properties of this substance, including its absorption spectrum, however, was sufficient to lead Kennaway to suspect that it might be a carcinogen. This indeed proved to be true.

In the search for this substance, or something like it, in pitch, Hieger extracted from several tons of this material a small quantity, not of 1:2:5:6-dibenzanthracene as might have been anticipated, but of a somewhat related compound which Cook, Hewett and Hieger (1933) showed to be 3:4-benzpyrene[1] by synthesizing the substance. This substance was also shown to be a powerful carcinogen for the skin and connective tissue of mice.

Conversion of Normal to Malignant Cell

Since the early 1930's when these discoveries were made, a considerable number of new pure synthetic chemical carcinogens has been prepared, and there seems no reason to suppose that we have by any means reached finality. Already some hundreds of substances are known to be carcinogenic for animals; but with this increasing knowledge comes less certainty as to how far these substances are concerned in the induction of human tumours. One might have supposed that with pure substances,

[1] Described at first as 1 2-benzpyrene.

that 3:4-benzpyrene is apparently not carcinogenic for rhesus monkeys in experiments where many tissues were exposed to the action of this substance for a period of up to ten years. It is clear therefore that one cannot grade substances in order of carcinogenic potency unless one also states for what animal and for what tissue. Clearly, for an effect to be produced, some work must be done, and whatever the mechanism involved, there must be some transfer of energy between the carcinogenic molecule and the cell that is in the course of being converted from a

gens may contain a clue to some physico-chemical properties shared by these substances, though not limited to carcinogens. Possibly in some energy transference of very limited range may be found the common factor between the chemical carcinogens and the physical carcinogens such as, for example, exposure to ultraviolet rays or to the ionizing radiations of X-rays or from naturally occurring radioactive substances, all of which are known to be carcinogenic both for man and for experimental animals.

The chemical and physical carcinogens at least show this in common, that although they initiate the process they certainly cannot be carried along from one cancer cell to the next in series. This is particularly obvious in the case of exposure to radiations. Moreover, between the cessation of exposure and the appearance of a recognizable cancer cell there must be a considerable time lag during which the cell need not be exposed any longer to the initiating energy source. What then maintains the character of the cell for an indefinite period thereafter?

Filterable Agents

Several theories have been advanced, two of which are certainly worthy of consideration. One of these postulates that the carcinogen acts in some way to render the cell susceptible to invasion by a specific virus which thereafter reproduces within the cell and so acts as the continuing cause of cancer. This theory has little to support it apart from a limited group of tumours in

1941) have shown rather similar effects in the ears of rabbits. Thus it seems that cells may be potentially malignant and lie dormant until stimulated. Perhaps this is not surprising when we consider that the cells of definitely malignant tumours can lie dormant for twenty years in man, only to recur after such a long latent period.

In seeking to disclose the essential stages in the conversion of a normal cell to a malignant cell, we thus have to deal with a mass of conflicting evidence. Many of the cells in the immediate neighbourhood of a carcinogen we know will not become malignant. Only perhaps a very small proportion of the cells exposed to the action of these agents will in fact respond in this specific way. Even when a tumour has been formed, we can be reasonably certain that not all the cells of the tumour are equally capable of continuing to multiply and maintain the malignant character of the growth. A little consideration of the size of the tumour that would be involved if every tumour cell were able to multiply indefinitely will show that most of them are destined to die. Indeed, the very characteristics that allow us to make a diagnosis of cancer are I believe deceptive in that they are largely the appearances of cells which are not in fact responsible for the maintenance of the tumour, but rather those that are undergoing degenerative changes to which the cancer cell is particularly liable, or differentiation, e.g. keratinization in squamous carcinoma.

Much ingenuity has been displayed, particularly by Daudel and the Pullmans (Daudel and Daudel, 1950; Pullman, Berthier and Pullman, 1950), on the mathematical considerations concerning the electrical charges carried by carcinogenic hydrocarbons, and they have drawn certain conclusions about the electrons distributed around a molecule; in particular the π-electrons in the meso-positions. All such considerations, however, must be considered against the background of the observed facts. Methylcholanthrene is apparently a more potent carcinogen for mice than 3:4-benzpyrene or 1 2·5:6-dibenzanthracene yet in fowls we have found that 1·2:5·6-dibenzanthracene will induce sarcoma more readily than either of the other two substances named. Moreover, Pfeiffer and Allen (1948) have shown

for such tumours the continuing cause and that loss of the virus for any reason might be followed by a reversion to normal behaviour by the previously malignant cells. While it is comparatively easy to imagine a single cell or a few cells undergoing a specific mutation which would cause them to behave in the way that we describe as malignant, it is very difficult to imagine the simultaneous mutation back to normal or to a lethal mutation of all the cells of a tumour. Thus the somatic mutation theory is fully compatible with the observed behaviour of the great majority of malignant tumours; but this of course does not by any means prove that the theory is right.

There can be no doubt that cancer is composed of a growth of cells previously normal which have undergone some peculiar kind of change. Whether it is necessary to regard this as a specific disease or whether it should be regarded merely as a special kind of tissue reaction which can be evoked by any one of quite a large number of agencies, is a much debated question.

REFERENCES

ANDERSON, W. (1950). *Acta unio internat contra Cancrum*, 7, 41

BERENBLUM, I (1941) *Cancer Research*, 1, 44.

CLAR, E. (1929). Quoted by Haddow and Kon, *Brit. med Bull.* 1946–7, 4, 314.

COOK, J. W., HEWETT, C L. and HIEGER, I. (1933). *J. Chem Soc* p. 395.

DAUDEL, P. and DAUDEL, R (1950) *Acta unio internat. contra Cancrum*, 7, 91.

DEELMAN, H. T. (1924). Quoted by Berenblum, *Brit. med Bull.* 1946–7, 4, 343.

HANAU, A. N. (1889). *Fortschr. Med.* 7, 321

HIEGER, I (1930). *Biochem. J* 24, 505.

HUEPER, W. C, WILEY, F. H and WOLFE, H D. (1938). *J industr Hyg* 20, 46.

KENNAWAY, E L. (1924). *J Path. Bact* 27, 233.

KENNAWAY, E. L. and HIEGER, I. (1930). *Brit med J.* 1, 1044.

LEITCH, A. (1922). *Brit. med. J.* 2, 1104.

[illegible] ted by Cook and Hewett,

[illegible], Imperial Cancer Research Fund.

which cell-free agents can be demonstrated as causal factors. From these few examples some pathologists like to generalize and suppose the presence of similar viruses in all tumours. This, however, ignores the negative evidence which is considerable and has remained consistent over a long period of years. There cannot be any doubt that for certain tumours there is present in the cells a filterable agent which exactly reproduces in a susceptible host the type of tumour from which the agent has been extracted. For such tumours it is idle to argue that a virus is not concerned. Such arguments merely beg the question of the definition of a virus.

But it must be pointed out that there are already known several distinct viruses associated with different histological types of cancer in different species of animals, and that these viruses are distinct from each other. There can be no valid theory of a single viral cause for all cancers. The virus-induced tumours have certain curious features in common, which distinguish them in some ways from the general run of tumours. For example, in new hosts infected with the specific virus, most of the tumours will grow progressively and cause the death of the animal; but a small proportion of them will undergo spon-t[illegible] ir[illegible] [illegible] that virus-induced tumours are the exception rather than the rule.

Somatic Mutation

The other theory that seems to merit consideration is that the carcinogens and possibly some viruses can cause somatic mutations. This theory can hardly be put to experimental proof, but circumstantial evidence can be adduced in favour of it. Most of the carcinogenic agents are also known to be mutagenic agents, as tested on simpler organisms, and the continued behaviour of the tumour cells in an unusual way could be accounted for on the basis of their mutant origin.

The spontaneous retrogression of the virus-induced tumours already referred to would rather suggest that in the case of such tumours no mutation has in fact occurred, but that the virus is

VI

The Functional Significance of Connective Tissue

A. H. T. ROBB-SMITH

VICTORIAN ladies had a drawing-room pastime called skeletonizing; it consisted of macerating plants to leave the fibre skeleton as an elegant object. It was no new idea, for Frederick Ruysch had devised the method in the seventeenth century for his museum specimens, which were more remarkable for their aesthetic appeal than for their medical instruction. However, these fibre skeletons do emphasize how the pattern of the body is, in reality, the pattern of its fibres, and also the resistance of these fibres to destructive agents. We can examine

ancient Égyptians suffered.

From the time of Fernel until the early part of the last century, histology, and much of the theories of physiology and pathology were based on the fibres of the body, but with the introduction of the cell theory and cellular pathology, interest was centred on the cell and its components and, until recently, the nature of the spaces between the cells was to a large extent ignored. Like all generalizations, this is not entirely true, for Claud Bernard emphasized the importance of the extracellular fluid in his discussion of the milieu interne, while Sir Rudolph Peters, in his concept of the cytoskeleton, opened the way to a biochemical approach to cell organization, and laid

PASSEY, R. D. (1922). *Brit. med. J.* **2**, 1112.
PFEIFFER, C. A. and ALLEN, E. (1948). *Cancer Res.* **8**, 97.
POTT, P. (1775). *Chirurgical Observations*, London.
PULLINGER, B. D. (1940). *J. Path. Bact.* **50**, 463.
PULLMAN, A., BERTHIER, G. and PULLMAN, B. (1950). *Acta unio internat contra Cancrum*, **7**, 140.
ROUS, P. and KIDD, J. G. (1941). *J. exp. Med.* **73**, 365.
SCHAR, W. (1930). *Le Cancer*, **7**, 205.
TSUTSUI, H. (1918). Quoted by Haddow and Kon, *Brit. med. Bull.* 1946–7, **4**, 314.
YAMAGIWA, K. and ICHIKAWA, K. (1915). Quoted by Haddow and Kon, *ibid.*

The Nature of Connective Tissue

We may regard connective tissue as a continuous fluid matrix varying in consistency from the limpid Wharton's Jelly of the umbilical cord to the hardness of bone, in which is lying an *interlacing fabric of fibres of different sorts and is bathing cells* which may be isolated or closely set, orientated to form an integumentary covering, tubules for absorption, secretion or excretion, the transmission of oxygen, blood or lymph or specially patterned as muscles or nervous tissue. Naturally we cannot ignore the cells; *Virchow's dictum may have been over-narrow* and Needham has pointed out that we must regard proteins, in a constant ebb and flow, as the responsible structures in living architecture and metabolism, yet it is the cell and its activity which ultimately determines the pattern of the extra-cellular matrix.

The character and thickness of the connective tissue layer varies from organ to organ. In the lungs and endocrine glands it is extremely thin except for the capsule; in the liver and kidneys it is somewhat thicker and in the heart and muscular organs it forms a significant portion of the viscus; this is also true of the intestinal tract and the skin, while the bones, joints and tendons are formed entirely of connective tissue, but it is important to remember that the connective tissue is continuous throughout the body.

It is really immaterial which organ we select to study the connective tissue in greater detail, but I shall choose skin, though it will be necessary to discuss certain aspects of other organs.

A section of skin examined at moderate magnification reveals

respects; for instance the blood supply appears very sparse, whereas a thick injected section of the same sort of region reveals

great stress on the physiochemical activity of the cell surface and interfacial tension.

The recent change in viewpoint with regard to connective tissue so that it has become a field of active research and clinical misconceptions, has stemmed from a number of sources. Textile chemists have been deeply interested in the structure of natural and artificial fibres for years and it was fortunate that this work was being done by men like Astbury, Schmitt, Highberger and many others, who were able to see beyond their immediate problems and develop the science of molecular biology (1), and last term Professor Randall presented this aspect of connective tissue to you.

In the field of human pathology, Klemperer and his colleagues at the Mount Sinai Hospital, New York, had been interested in disseminate lupus erythematosus for many years and in 1942 they published a paper entitled 'Diffuse Collagen Disease' (2) in which they pointed out that in scleroderma and disseminated lupus there were conspicuous and systemic alterations in the extra-cellular components of the connective tissue and that from the point of view of morphogenesis they might be regarded as disorders of the connective tissue system, a reaffirmation of Bichat's concept of pathology. Comparatively little attention was paid to this masterly paper, until the recognition that cortisone appeared to be effective in the treatment of many disorders of connective tissue such as rheumatoid arthritis, disseminate lupus, etc., when as Klemperer (3) said, 'the impatience of clinical investigators and a peculiar worship of diagnostic terms have led to an exaggerated popularity of the diagnosis of Collagen disease. There is a danger that it may become a catch-all term for maladies with puzzling clinical and anatomical features It is not a term applicable to diagnosis and certainly does not define the morbid process of the diseases grouped together.'

I would like to try and bring together this evening some ideas on the function both normal and abnormal of connective tissue, but it would be well first of all to consider its anatomy.

molecular level and I think it might be useful to try and realize the orders of magnitude with which we are dealing for it is all too common for confusion to arise owing to lack of clear thinking on this point. The figure (Plate I, Fig. 1[1]) shows collagen fibres in the light microscope, in the electron microscope and schematic diagrams of its molecular pattern. If one could visualize the man from whom the fibres were taken at the same magnification as the illustrations, at the level of the light microscope (Fig. 1(*i*)) he would appear just under a mile tall; at the maximum resolution of the electron microscope (Fig. 1(*iii*)) he would appear 70 miles tall, about the distance from London to Dover; but the magnification of molecular pattern (Fig. 1(*vi*)) is a thousand times greater so that he would appear 70,000 miles tall, ten times the diameter of the earth.

The nature of reticulin and its relationship to collagen has been a matter for discussion for the last fifty years or so, but I believe that Harold Kramer (5), working in my laboratory, has cleared away much of the confusion by the examination with the electron microscope, X-ray diffraction, etc., of material which histologically consisted of reticulin almost free from histological collagen. It would seem that reticulin consists of a network of very fine disorientated fibrils of about 100 Å in diameter but showing the 640 Å periodicity of collagen; this network is lying in a membrane which appears to consist of polysaccharide and non-fibrillary collagen, probably only partially orientated.

The nature of elastica is at the present time somewhat obscure. Chemically it is a polypeptide with a high content of glycine, alanine and valine but a very low content of polar residues and X-ray diffraction would suggest that the molecules are largely disorientated, as is the case with rubber, which would be in keeping with its elastic properties.

The earliest electron micrographs were made by Wolpers (6) who found branching fibres lying in a membrane, then Gross (7) prepared elastica and found it to consist of fine twisted threads lying in an amorphous matrix, but subsequently he and others found that these threads were probably artefacts derived from the trypsin solution and were not elastica at all (8).

[1] This plate will be found at the end of the book

component of connective tissue, and the reaction under polarized light indicates that these bundles are made up of orientated fibrils. A section stained by orcein reveals the network of elastic fibres, another of the fibrillary components of connective tissue.

If we examine the dermis under higher magnification, when impregnated with silver we find that the interlacing pattern of dense bundles of collagen which were birefringent under polarized light stain a light golden brown and are not so conspicuous as the finer black fibrils, so-called reticulin, and we can see how these are condensed at the dermic epidermic junction and around the blood vessels—the basement membrane. Now if we treated the section with periodic acid followed by Schiff's reagent—a method known as the Hotchkiss or McManus stain which, with certain limitations and provisos which I will not go into, will reveal the presence of carbohydrate (14)—we find that the spaces between the fibres will stain faintly, while around the reticulin fibrils it will stain a more intense red. So we are beginning to get an idea of the nature of connective tissue—it consists of an interlacing network of collagen, reticulin and elastica fibres lying in a non-fibrillary matrix of ground substance which has a variable carbohydrate content.

If we examine the coarse birefringent collagen bundles with the electron microscope, they have a characteristic appearance with cross striations having a periodicity of 640 Å; a great deal of work has been carried out in relation to the structure of collagen fibre, trying to determine in the light of the characters that can be displayed by means of X-ray diffraction, electron microscopy and physico-chemical analysis, just how the polypeptide chains are arranged; some favour a sheet-type structure, others a cylindrical or spiral molecule, but much more work is needed before the picture is clarified (4); from the point of view of chemistry we can state with reasonable certainty that collagen is a polypeptide, in which a third of its residue consists of the amino-acids proline and hydroxy proline and a third of glycine together with a small amount of carbohydrate. *In discussing* the nature of collagen we have glibly moved from the magnification of the light microscope away beyond the limits of the electron microscope and the X-ray diffraction beam to the

in metabolic function but the idea that it is a stagnant pool of salt water whose character is determined by osmosis must be abandoned. It is a highly dynamic fluid and work with radio isotopes has emphasized the rapidity of electrolyte dispersal and exchange.

Under the microscope the ground substance appears optically homogenous, indeed its very existence was questioned for many years until Sylvia Bensley (13) displayed it by some ingeniously simple experiments and its chemical constitution can only be determined by rather fierce extraction methods; the histochemical methods available are also relatively crude (14), the Hotchkiss or Periodic Acid Schiff reaction merely indicating the probable presence of polysaccharides, and under certain abnormal conditions ground substance may show the phenom-

droitin sulphate or heparin, the non-acetylated sylphuric ester. There is another method of investigating the changes in ground substance by the use of the enzyme hyaluronidase (86)—the spreading factor of Duran-Reynals—which hydrolyses both hyaluronic acid and chondroitin sulphate. It is possible to examine tissues after exposure to the enzyme and by noting the differences, deduce the sites of mucopolysaccharides; but once again it is not over-precise, nor does the electron microscope help, and so we must work by the correlation of numerous approaches in order to assess changes in ground substance.

To recapitulate, we can define connective tissue as tissue fluid in a matrix of mucopolysaccharides and scleroproteins, in varying degrees of polymerization but largely disorientated, in which are lying orientated fibres of varying thickness of collagen and elastica.

There are structural variants of connective tissue; we have the limiting membrane of muscle, the basement membrane of gland tubules and Disse's space around the hepatic cords which consist of a thin membrane of polymerized mucoprotein in which fine collagen fibrils are coursing; then we have tendon where

Recently Hall, Reed, and Tunbridge (9) have described elastica as made up of two components—ill-defined, branching threads lying in an amorphous matrix, the proportion of the two elements differing from one site to another with a variable content of sulphate polysaccharide in combination with the protein, the non-fibrillary material having the higher polysaccharide content. Lansing (10) has obtained similar results with the electron microscope but the latest paper by Tunbridge (11) and his colleagues suggests that certain connective tissue fibres which stain with the light microscope as elastica, under the electron microscope have the characters of collagen. It looks as though we are going to be up against the same sort of problem that there has been with reticulin and that it will have to be solved in an analogous manner.

Lastly what is the nature of the ground substance; well, in part it consists of tissue fluid, derived largely from the blood plasma, and one should remember that the extravascular albumen pool is twice the mass of that within the vessels; but in addition it has a high content of mucopolysaccharides, chiefly hyaluronic acid and chondroitin sulphate and glyco-proteins (which are mucopolysaccharide-protein complexes) in a varying degree of polymerization; ordinarily it has a gel-like consistency and the histological evidence, for what it is worth, would suggest that the glycoproteins are most highly polymerized and therefore less soluble at the basement membrane where it is in close relationship to the reticulin and collagen fibres, but in the inter-fibrillary spaces it is less condensed and may, under abnormal conditions, become depolymerized, more fluid and finally water soluble.

Cowdry (12) has summarized the essential differences between ground substance and blood plasma as follows:—It has less diffusable organic solutes, it contains mucopolysaccharides and scleroproteins, a higher concentration of hormones, more lactic acid and mineral salts, a greater range of fluidity and hydrogen ion concentration and a variable amount of water, salt and glucose depending on storage and release from storage. I would be unfitted to discuss the general metabolic significance of the extra-cellular fluid and the changes that may occur in it

with depolymerization of mucopolysaccharide and a phosphorylative glycogenolysis enzyme cycle proceeds in which the alkaline phosphatase probably acts as a transphosphorylase (18). The resultant is the formation of a phosphate calcium acceptor in the ground substance or cartilaginous matrix and calcification commences. There is much dispute as to the stages that take place in calcification but it is generally believed that bicalcium phosphate is first formed and then is transformed to tricalcium phosphate, calcium carbonate and apatites (19). The ultimate formation of bone with its patterned arrangement of collagen and elastic fibres in relation to the bone salt crystals which are about 400 Å long and 25-50 Å wide and the appearance of haversian systems is intimately related to the development of a vascular pattern (20).

Formation of Connective Tissue

So much for the morphology and nature of connective tissue; now how is it formed, and the answer, as is so often the case in biology, is that we do not really know. It is obvious that under ordinary circumstances the connective tissue cells that we call fibroblasts are the cytogenic foci and originally it was believed that fibrin fibres, as extensions of the fibroblast cytoplasm, were converted into collagen but now this can be definitely excluded (21), although it may well be that when connective tissue fibres are forming, fibrin, by contraction, plays a part in the orientation and alignment in a purely mechanical way (22). The embryonic mesenchyma is rich in mucopolysaccharides but is relatively acellular and in this, collagen-type fibres appear. Then in tissue-culture experiments it was shown by Doljanski and Roulet (23) that the formation of collagen fibres could occur in a portion of a fibroblast culture isolated from the cells, and they suggested that the fibroblasts secreted an enzyme, which acted on a substrate, contained in the plasma and embryonic extract of the culture medium. More recently Keith Porter (24) has studied tissue cultures of fibroblasts with the electron microscope and has found that although cross banded collagen fibres are formed, yet these often have a different periodicity from mature collagen fibres, some having a period of 210 Å and

there is a high proportion of coarse collagen fibre with a variable admixture of elastica, embedded in a mucopolysaccharide matrix.

In the formation of cartilage there is an increased metachromacy of the ground substance indicating a higher content of chondroitin sulphate, while the coarse collagen fibres become more widely separated, and the mesenchymal cells divide, become swollen, with a high glycogen content in the cytoplasm and then become grouped but widely separated in the highly polymerized ground substance which is cartilage. It is rich in orientated argyrophil fibrils resembling reticulin except that they show little branching but their patterning is significant in relation to the shaping of cartilage. In addition there are collagen and elastic fibres and in fibro- and elastic cartilage they are in higher proportion. Precise chemical analysis of cartilage is lacking; it contains about 70 per cent water and of the dried residue, 80 per cent consists of collagen and chondroitin sulphate in varying proportions according to the site from which it is obtained (15); there is good evidence that much of the mucopolysaccharide is combined with collagen in the form of glycoproteins.

It should be appreciated that a joint is merely an area of connective tissue free from cells and fibres in which the ground substance—in the form of synovial fluid—consists of tissue fluids with a high content of hyaluronic acid. Edlund (16), using the manometric method evolved by McMaster (17) showed that the tissue resistance of a joint cavity was identical with that of loose connective tissue and in the formation of bursae and ganglia we have the development of miniature joints in connective tissue.

The formation of bone is essentially a patterned calcification of ground substance associated with increased vascularization, and although as a histologist one distinguishes bone formation in fibre, by apposition and enchondral ossification, yet the changes in all three are really the same; there is an increase of mucopolysaccharides in the ground substance with a separation of the collagen fibres and a morphological modification of the connective tissue cells; then there is a change to an alkaline pH

with depolymerization of mucopolysaccharide and a phosphorylative glycogenolysis enzyme cycle proceeds in which the alkaline phosphatase probably acts as a transphosphorylase (18). The resultant is the formation of a phosphate calcium acceptor in the ground substance or cartilaginous matrix and calcification commences. There is much dispute as to the stages that take place in calcification but it is generally believed that bicalcium phosphate is first formed and then is transformed to tricalcium phosphate, calcium carbonate and apatites (19). The ultimate formation of bone with its patterned arrangement of collagen and elastic fibres in relation to the bone salt crystals which are about 400 Å long and 25-50 Å wide and the appearance of haversian systems is intimately related to the development of a vascular pattern (20).

Formation of Connective Tissue

So much for the morphology and nature of connective tissue; now how is it formed, and the answer, as is so often the case in biology, is that we do not really know. It is obvious that under ordinary circumstances the connective tissue cells that we call fibroblasts are the cytogenic foci and originally it was believed that fibrin fibres, as extensions of the fibroblast cytoplasm, were converted into collagen but now this can be definitely excluded (21), although it may well be that when connective tissue fibres are forming, fibrin, by contraction, plays a part in the orientation and alignment in a purely mechanical way (22). The embryonic mesenchyma is rich in mucopolysaccharides but is relatively acellular and in this, collagen-type fibres appear. Then in tissue-culture experiments it was shown by Doljanski and Roulet (23) that the formation of collagen fibres could occur in a portion of a fibroblast culture isolated from the cells, and they suggested that the fibroblasts secreted an enzyme, which acted on a substrate, contained in the plasma and embryonic extract of the culture medium More recently Keith Porter (24) has studied tissue cultures of fibroblasts with the electron microscope and has found that although cross banded collagen fibres are formed, yet these often have a different periodicity from mature collagen fibres, some having a period of 210 Å and

occasionally fibres with a spacing of 1,100 Å are observed, and Wyckoff (25) has a beautiful electron micrograph of a fibroblast apparently with a fringe of short collagen fibres. Now it is known that in embryonic or growing connective tissue there is often a high proportion of short collagen fibres with tapered ends and these may have a cross striated period of 210 Å instead of the 640 Å seen in mature collagen.

Another aspect of work on the formation of collagen has arisen from studies of the material first isolated by a group of Russian scientists and called by them procollagen (26); it is obtained by dialysis of a citric acid extract of skin, and on examining this material with the electron microscope, it is found to contain a mixture of the normal 640 Å collagen fibres and fibres with a very long spacing 2,000–3,000 Å. If this procollagen is purified, no fibrillary collagen is formed, and it occurred to

they produced varying proportions of long spacing or normal spacing collagen fibres; on the other hand polysaccharides, such as hyaluronic acid or heparin or chondroitin sulphate, did not induce fibre formation and they concluded that a carbohydrate fraction was essential for fibre formation and that certain complex chemical substances of a non-specific nature had to be present to induce precipitation and alignment of the protofibrils. Schmitt and his colleagues (28) have recently described a third form of particle, a polarized asymmetrical type of long-spacing segment, which can be precipitated from collagen 'solutions' by the action of adenosine triphosphoric acid; further they have shown that the three structural forms of collagen are interconvertible under suitable conditions.

Harkness (29) has found by isotope studies that in the initial stage of fibre formation, there is formation of non-fibrous soluble protein analogous to the Russians' procollagen and gradually the soluble procollagen-carbohydrate complex becomes dissociated during conversion into the insoluble mature collagen fibre. In fibroblast tissue-cultures, Mancini and de Lustig (30) have found that hyaluronidase inhibits the formation

of collagen fibres and we know that this enzyme hydrolyses certain mucopolysaccharides.

Now, Albert Fischer (31) has put forward a hypothesis, supported by experimental work, that the plasma proteins form a readily available source of amino-acids for the synthesis of complex cellular and extra-cellular protein, that the plasma proteins are broken down by proteolytic enzymes of the fibrinolysin or plasma trypsin type or else are fixed on to cells by a mechanism analogous to that occurring in blood coagulation and so are readily accessible to cellular proteolytic enzymes.

One of the objections to this hypothesis of transpeptidisation catalysed by proteolytic enzymes is that enzymes of this type also catalyse the hydrolysis of the peptides and the competition between water and the amine for the carboxyl group will almost certainly be decided in favour of water, as the hydrolysis equilibrium of all peptides is far to the side of complete hydrolysis; however, in connective tissue the mucopolysaccharides are very active in water binding and so water might not enter in competition and thus transpeptidization might proceed. Haurowitz (32) has put forward a detailed hypothesis of protein synthesis as a two-phase reaction, involving alignment of amino-acids as a peptide chain which is then folded to form a three-dimensional globular protein molecule.

Thus we have good evidence that mucopolysaccharides play an essential part in collagen fibrillogenesis and that the fibroblast at least potentiates this, but it is clear that other factors are also involved in the formation of normal connective tissue For instance, in the healing of wounds in scurvy (33) there is a deficiency of collagen and reticulin formation and the ground substance contains a gelatinous fluid which is not metachromatic but has a high protein content, there is a deficiency of phosphatase in the fibroblasts, which have the cytological characters associated with active formation of cytoplasmic protein and an increase of glycoprotein in the serum (34). On exhibiting ascorbic acid, the phosphatase content of the fibroblasts increases, the ground substance shows metachromasia, there is formation of reticulin and collagen, and serum glycoproteins fall. One can deduce from this that ascorbic acid assists in the

polymerization of mucopolysaccharides and that this is necessary for collagen fibril formation, although it may be that the scorbutic fibroblasts are only able to form a globular rather than a fibrillary scleroprotein of the collagen type; it is uncertain whether the phosphatase deficiency is *post hoc* or *propter hoc* but it is certain that phosphatase plays an essential part in carbohydrate metabolism and there is deficient phosphorylation in scurvy; as to whether phosphatase plays any direct part in protein synthesis is somewhat doubtful (35). So we have further evidence from ascorbic acid deficiency of the interrelationship between the mucopolysaccharides and collagen fibre formation. Kramer (36) has put forward the hypothesis that mucopolysaccharides might act as an anionic detergent and by breaking down linkages in the globular form of collagen, which has been synthesized through the medium of the fibroblast, would promote curling and some degree of orientation of the collagen molecules, and that the collagen protofibrils crystallize as the carbohydrate moiety is lost by dissociation, due to a change in the local electrolyte concentration or enzyme activity. This hypothesis is in keeping with the facts of fibrillogenesis as far as they are known, conforms to known processes of synthetic fibre formation, and would explain the formation of collagen in sites relatively isolated from cellular activity as is often seen in pathological changes. It is of some interest that Mercer (37) has put forward an analogous mechanism for the biosynthesis of silk fibrosin.

Two further questions remain with regard to the formation of connective tissue, the source of elastica and of the mucopolysaccharide. Elastica formation is an enigma (38). In the embryo and in healing wounds it appears much later than the collagen fibrils, and indeed it would appear that collagen is a prerequisite for its formation. In tissue culture the fibres are produced only by tissues which contain them *in vivo*; in heart muscle cultures they are most dense when subjected to the pull of the

[illegible]
chromatic material which takes on a beaded form and then fine

non-argyrophil metachromatic fibres appear which gradually develop the staining reactions of elastica; at first the fibres are irregularly arranged in a fine meshwork and then develop into the fine branching and coiled refractile threads which are characteristic of elastica. Hass (40) has suggested, on the basis of some interesting experiments, that it might be formed as a fibrillary membrane at lipoid aqueous interspaces in the tissues and such a mechanism would be in keeping with observations such as those of Montgomery (41) on the healing of experimental wound of the lung. We know from Lansing's (10) work that the mature fibres are associated with lipoid and Tunbridge and his co-workers (11) have shown an association with mucopolysaccharides, but the cytogenesis is very obscure.

With regard to the mucopolysaccharides of the ground substance, there are rival claimants; the presence in fibroblasts of granules and vacuoles having the histochemical reactions of mucopolysaccharides and the occurrence of large amounts of mucopolysaccharides in certain types of fibrosarcoma would appear to favour this cell type (42); on the other hand the mast cell is known to store, if not to synthesize, heparin—a non-acetylated sulphuric ester of glucuronic acid—and Asboe-Hansen (43) has put forward much evidence to suggest that it is also the source of the connective tissue mucopolysaccharides. I believe that the matter could perhaps be resolved as the result of a detailed cytological study of the Sanarelli myxoma virus and the mast cell reaction in scurvy, but it is interesting to consider that not so very long ago the function of the mast cell was completely unknown and now it is credited with the formation of heparin, connective tissue mucopolysaccharides, and, in the light of Riley and West's (44) work, histamine as well.

Turning from biosynthesis to metabolism, we have even less information. There have been studies of collagen by Neuberger (29) using C^{14} labelled glycine which would indicate that in the growing animal, the amino-acid is taken up by the developing fibres and to a large extent retained there; in the adult animal very little is taken up. This, together with studies on scorbutic collagen formation, would suggest that there is very little katabolism but a constant anabolism (45), but as

Neuberger rightly points out, one must be very cautious in the interpretation of isotope metabolic studies. With regard to the ground substance we are even more ignorant; there is indirect evidence, to which I will refer in a minute, that there is depolymerization and polymerization of glycoproteins under suitable stimuli, with release of glycoproteins into the blood and urine, but further than that we cannot go.

The Functions of Connective Tissue

So far I have tried to provide you with a picture of the nature of connective tissue and how it is formed. When we consider its function we can look at it structurally, metabolically and in terms of disordered function or pathology. Its mechanical function is generally accepted, though it inevitably leads us into teleology. Sir D'Arcy Thompson (46) studied the adaptation of the internal structure of bone to the stresses and strains to which it is subject and it has been shown that the tensile strength of bone is of the same order as cast iron; cartilage provides a tissue which combines a certain amount of rigidity with considerable flexibility and resilience, and Matthews (47) has recently suggested that there is a higher ratio of mucopolysaccharide to collagen in weight-bearing joints. Collagen has great tensile strength with flexibility but negligible extensibility and so in tendons transmits, with a minimum loss of power, muscular contraction to the bones, and elsewhere maintains the shape of organs, and in its looser forms combined with ground substance forms an admirable flexible packing material. When loose collagen is combined with elastica it provides a tissue with great elasticity and resilience as in blood vessels and the dermis.

A major function of connective tissue is the repair of physical trauma after injury or the restoration of tissue after inflammatory reactions; however, new-formed collagenous tissue contracts, and without elastic tissue will stretch under tensile strain and much of the chronic diseases and dysfunctions of the body are the result of this hepatic or renal failure due to a scarred liver or kidney; cardiac failure due to mitral stenosis; intestinal obstruction due to a scirrhous carcinoma, an aneurysm of the aorta; the initial pathogenesis of these and many

other conditions I could mention may be due to cellular dysfunction, but it is the collagen that strangles the organ or allows it to burst.

However, we often ignore the metabolic function of connective tissue because we tend only to think of intracellular metabolism or general metabolism, but it should be realized that all metabolic processes, if transmitted, are transmitted and therefore modified by connective tissue surrounded by cell surfaces in which much of their energy is concentrated; connective tissue has an intercommunicating system of protein fibrils whose energy surface is immense and if we consider the speed of transmission of drugs and metabolites and their control, we can gain some inkling of its importance in physiology and functional pathology, and I would like to discuss this aspect for a few minutes.

It is well known that if we inject an animal with a dye such as trypan blue and examine the tissues in a few days, the dye will chiefly be collected in the cells of the reticulo-endothelial system, yet the animal becomes blue in a matter of minutes; King (48) has shown that at this stage, the dye is not in cells but is bound by connective tissue fibres all over the body, in collagen, reticulin and especially elastica, and that it is not absorbed into cells until 24 hours or more have elapsed; Barcroft (49) made similar observations with regard to the Wharton's jelly of the umbilical cord and suggested that it might serve as a non-vascular conducting path of metabolites from the placenta to the embryo and a similar mechanism might apply in the nutrition of cartilage and the cornea.

The ground substance is hydrophilic (50) and may store water and electrolytes, and it is almost certain that under normal conditions there is no free fluid but the electrolytes move through bound water (51). Opie and Rothbard (52) have shown that connective tissue takes up water by hydration as does gelatin and not by osmosis as is the case of parenchymatous tissue. The extensibility of connective tissue in the living organism depends on its content of water and when water content rises extensibility increases (53). In relation to the haemato-parenchymal barrier, the ground substance and fibrils are

affected in antagonistic fashion by many chemical substances, thereby enhancing the regulator action of this barrier (54).

The storage and release of calcium and other metals in bone is generally accepted, but it is not always appreciated how rapid the resorption and reformation of bony stroma can be; for instance [illegible]
culae [illegible]
reapp[illegible]
say in [illegible]
can act directly on bone and the initial phase is a depolymerization of ground substance of the bone with a rise of serum glycoproteins, *the depolymerization results in a reduction of mineral* binding capacity and calcium and phosphate are liberated; there is an increased excretion of glycoprotein in the urine and *on occasion mucoprotein phosphate carbonate casts are formed*; these experiments, in addition to clarifying many of the obscurities of bone resorption, may have some bearing on Butt's (57) claims that hyaluronidase is of value in preventing calculous formation.

INFLUENCE OF HORMONES

There is other evidence with regard to hormonal influence on connective tissue, whether one considers endocrine disorders which one may look on as caricatures of normal hormonal balance, or experimental procedures. In thyroid deficiency, the myxoedema is associated with a considerable increase of ground substance which is metachromatic, indicating some degree of depolymerization and an increase of tissue mast cells and water retention (58); all these effects are reversed on thyroid medication but it should be appreciated that there are certain forms of nodular myxoedema in which there is no evidence of hypothyroidism although the histology is identical (59) and it may well be that the water retention is not a primary effect of thyroid deficiency but is consequent on the altered metabolism of ground substance; furthermore it may be that the changes in connective tissue in myxoedema may not be due to a deficiency of thyroxin, but an excess of thyrotrophic pituitary hormone; *Asboe-Hansen and others* (60) have shown that both in malignant exophthalmos and pretibial myxoedema, which is com-

monly associated with a hyperthyroid state, similar tissue changes are seen to those in hypothyroid myxoedema and this can be induced experimentally by treatment with pituitary thyrotropic hormone; this raises the problem of focal reaction of connective tissue to hormonal stimulation which is particularly well seen in regard to the sex hormones.

The alterations in the stroma of the breast and uterus in pregnancy are well known, and though the connective tissue changes may be secondary to the epithelial or muscular proliferations which are directly hormonally stimulated, yet there is experimental evidence that there is a direct effect as well (61); in regard to the uterus, Harkness (62) has shown that there is an actual increase of collagen content during pregnancy and this is reversible, returning almost to normal within a week of parturition. Then there are the Lipschutz (63) type of experiments in which fibromatosis occurs not only in the uterus, but elsewhere in spayed animals treated with large doses of oestrogens. Elliott and I (64) did some experiments of this type on ascorbic-deficient animals and found that the fibromatosis occurred even though the animals died of scurvy; indeed they died earlier than the untreated controls, which might suggest that the hormonal induced fibrous proliferation utilized the limited amount of residual ascorbic acid. Even more localized reactions are the changes in the symphysis pubis induced by the luteoid hormone relaxin (65) in which there is absorption of bone and increase of ground substance, while in the ovary, a rhythmic remodelling of connective tissue associated with the maturation of graafian follicles at each oestrus cycle (66) Analogous changes, though more general in distribution, are the swelling of the sexual skin in monkeys which makes the mandrill such a colourful beast (67) and the testerone swelling of the cock's comb (68). It would seem that oestrogens and testosterone induce synthesis of mucopolysaccharides and increase their state of hydration (69), whereas gonadotrophins induce depolymerization with increased permeability (70), an effect analogous to that of hyaluronidase

The observation that there was impaired healing of wounds in cases under cortisone therapy (71) stimulated work on the

effect of adrenal cortical hormones on connective tissue, with the consequence that there has been considerable confusion; it was not recognized that there are marked differences in response in different animal species (72) and that the guinea pig, that most obliging of experimental animals, was unhappily extremely resistant to cortisone (73).

In general it may be said that the effects of cortisone on connective tissue support the views expressed by Albright in 1942 (74), seven years before Hench and Kendal (75) introduced the wonder-drug to clinical medicine, that is to say that its action is antianabolic rather than katabolic. Thus we find that there is an increase of free hydroxyproline (76), a failure of incorporation of sulphate into chondroitin sulphate (77), a failure of fibroblast formation or proliferation, but where fibroblasts are present, collagen and ground substances are normally laid down and immature connective tissue becomes mature. Mast cells lose their granules and appear degenerate (78) and this is even true of mast cell tumours of dogs (79). On the other hand, desoxycorticosterone (80) has an opposite effect with stimulation of fibroplasia and increase of metachromatic ground substance, while adrenalectomy, although not affecting collagen fibres or their water absorption, yet the ground substance is unable to retain water, allowing a shift in the balance of cellular and extra-cellular compartments.

In considering the changes in water and electrolytes, the question must remain open as to whether the hormones directly modify this aspect of metabolism, or induce changes in the formation and degree of polymerization of the mucopolysaccharides and scleroproteins, with consequential changes in bound water and electrolyte mobility of a purely physiochemical nature (82).

It is logical now to consider the changes in connective tissue in ageing. Ageing has been defined as colloidal dehydration (83) and this process is well seen in connective tissue, which in the embryo has a very high water content and gradually dries up, with increasing years; as this occurs the side chains of the polypeptides are drawn together with loss of physical response and chemical activity; in the stable polymerized state there is

an increased deposition of calcium, well seen in the aorta and in ageing cartilage, and the condensation of the connective tissue may be a factor in limiting the size of an organ (84), for if the cellular environment becomes resistant to growth, the cells may become confined and stop expansive growth, the reduction in chemical reactivity of connective tissue will limit the transport of metabolites to cells and removal of breakdown products, and so there will be a slowing of cellular metabolism and the biological machine will run down. The morphological counterpart to this hypothesis can be seen in the thickening basement membranes of ageing organs.

PATHOLOGY OF CONNECTIVE TISSUE

I shall not deal with the congenital abnormalities of connective tissue, conditions such as epidermolysis bullosa, Ehlers-Danlos syndrome, elastosis dysplastica, fragilitas ossium, and so on, though they are just as likely to provide the clues to normal connective tissue formation, as Garrod found in inborn metabolic errors. Nor can I do more than touch on the pathology of connective tissue, but there are one or two aspects of this to which I would like to draw your attention.

Response to Injury and Infections

I have already mentioned the response to injury of connective tissue, but it is well to realize that in a parenchymatous organ such as the liver or kidney, one may have extensive cellular damage, but provided the connective tissue matrix has not been damaged, if cellular regeneration takes place, the organ will be restored to normality both morphologically and functionally, and furthermore even when there has been fibrosis and scarring, this will at least in part be resorbed. In relation to infections, one considers principally the cellular and humoral reaction, but the part that connective tissue plays should not be forgotten (85); the polysaccharide gel will, to some extent, immobilize micro-organisms, and though this may be hydrolysed by hyaluronidase-type enzymes formed by certain organisms, yet there is the presence of the non-specific hyaluronidase inhibitor which is probably heparin (86), even though we may not accept the

whole of Haas' concept of anti- and pro-invasins (87); next the coarse collagen membranes may encourage the localization of an infective process, though this may be broken down by the organism's proteases, and lastly with resolution and new formation of collagen, the infection may be isolated.

In the case of certain virus infections, such as influenza, the entry of the virus into the cell is associated with the occurrence of certain mucinases, probably having as their substrates, mucoitin rather than chondroitin sulphate; neverthless, the mucoprotein viral inhibitor described in the urine by Tamm and Horsfall, or something very like it, is present in connective tissue and may well be a factor in the body's viral defences (88).

In the normal immune response, antibodies will be present in the ground substance to react with parenterally introduced antigens, and it has been shown that the extravascular plasma protein pool is about twice the mass of that in the circulation (89). Hypersensitivity is dependent on the reaction of antigen with sessile antibody contained in or on cells and although we have little or no knowledge as to the proximate biochemical lesion that results, its effects on smooth muscle and capillary permeability are familiar enough; the ground substance becomes oedematous and in addition to the cellular reaction, there may be the development of fibrinoid change in the connective tissue (90)—the homogeneous, highly refractile, band-like material which histochemically would appear to be depolymerized mucopolysaccharide, possibly associated with exudation of fibrin. There is good evidence that there is a release of a histamine-like substance which has been presumed to be derived from the damaged cells. In the light of Riley and West's (44) work, one must wonder about the interpretation of Stuart's (91) observation that in anaphylaxis there is a rupture of mast cells with disappearance of the granules and that this change is minimal when antihistamine drugs are given; on the other hand S[illegible] causes a rapid in-c[illegible] of metachromatic g[illegible], the whole problem of the heparin-histamine mast cell story needs considerable investigation.

Apart from these specific pathological reactions of connective tissue, we have the non-specific reaction—Eppinger's serous inflammation (54, 93), which I have recently discussed elsewhere (94). Briefly this postulates that, as a result of damage to capillary endothelium, cell membranes or ground substance, there is an accumulation of a protein-rich fluid in the connective tissue spaces and the consequent separation of the capillary from the parenchyma impairs the transport of nutritive materials and cellular metabolites with a secondary disturbance of cellular function. This type of connective tissue change is reversible if the damaging agent is removed or is low in activity; however, if the serous exudation persists, there will be an irreversible change in the ground substance with the formation of scarring or hyaline change (95), or amyloid deposition if conditions are suitable; what these are is unknown, but it would seem that some of the underlying factors are ascorbic acid deficiency, associated with a high globulin content of the ground substance, which may be antibody globulin or consequent on a disturbance of protein metabolism (96).

One of the many problems in malignant disease is the mechanism of infiltration and the development rather than the transport of metastases; it is clearly the resultant of a reaction between the neoplastic cells and the connective tissue stroma. Coman (97) has suggested that malignant cells show a decreased adhesiveness, amoeboid movements and the presence of an enzyme of the hyaluronidase type, although this has not been confirmed and may well have been due to bacterial contamination; certainly neither he nor Simpson (98) was able to show any increased invasiveness if hyaluronidase was injected in the region of a tumour. Turning to the stroma, Cramer (99) and Orr (100) have shown that in the pre-malignant phase of experimental carcinogenesis, the connective tissue undergoes spectacular changes and these only occur if the substance used is in fact carcinogenic, it is not a simple reactive change to an irritant. They found that there is a marked increase of mast cells, increased elastic formation and a change in the collagen fibres which are chiefly of the reticulin type, with a minimal cellular reaction; when malignant change occurs, there are

breaks in the elastica and disappearance of mast cells. Bellingham, Orr and Woodhouse (101) have shown by transplantation experiments that it is the stroma which is the determining factor in certain types of experimental carcinomata. The next stage in malignancy, that of infiltration, has been studied by Gersh and Catchpole (102) and Sylven (103). They have found that in a rapidly-growing tumour there is marked depolymerization of the ground substance with increased serum glycoproteins, and the reverse is to be seen in slow-growing tumours where there is evidence of a fibrous reaction. The contrast between a scirrhous tissue reaction and the lack of tissue reaction in an encephaloid carcinoma is familiar to all, but the difficulty is that given these morphological observations, one is uncertain how to interpret them; is it *post hoc* or *propter hoc*? The answer may partially be resolved by the introduction of more critical methods for detecting depolymerization and proteolytic enzymes.

Disseminated Lupus Erythematosus

Although there is much more that I would like to tell you of the pathology of connective tissue—the significance of changes in atherosclerosis and aortic medial necrosis where one has an admirable example of depolymerization taking place in relation to all the elements of connective tissue (104), and the nature and possible mechanism of the changes in the ill-named collagen diseases (105), in conclusion I would like to say a few words about disseminate lupus erythematosus, not so much from the point of view of its special pathology, but because it is an example of a disorder characterized by lesions in connective tissue in which we are beginning to get some idea of pathogenesis.

Disseminate lupus is a rather uncommon disease (106), chiefly seen in adult women. The patient presents a variety of vague symptoms of which the most constant is irregular fever with malaise and weakness; there will be attacks of swelling of the joints and very frequently a rash, particularly on the malar region of the face which is red and scaly I shall not describe the clinical signs or laboratory findings, but ordinarily the disease runs an irregular course with exacerbations and remissions over months or years, and until recently was invariably fatal. At

post-mortem the changes are often not very spectacular; there will be the rash, enlarged lymph nodes, the thickening of joints, and some effusion in the serous cavities. The heart may show non-bacterial vegetations.

Microscopically the constant lesion found in varying degree throughout the connective tissue is that the ground substance is often metachromatic and collagen fibres are separated and blurred and stain the deep red colour characteristic of 'fibrinoid' and an analogous appearance is found in the basement membrane, particularly of the kidney, and there is often degenerative changes in the cells of the connective tissues with some lymphocytic increase (107). That was the position about four years ago and it was suggested that the disease was due to some infective agent or was allergic, whatever that might mean. Then, in 1948 Hargreaves (108), in the Mayo Clinic, observed that in the marrow in this disease there was often clumping of leucocytes and leucocytes which contained homogeneous inclusion bodies—so-called lupus erythematosus or L.E. cells. It was then found that it was possible to induce similar changes in normal marrow cells by adding the serum from a patient with the disease and the factor—the so-called L.E. factor—was present in the globulin fraction and was antigenic and only occurred in this disease (109); it was shown that the L.E. cell was a neutrophil leucocyte or macrophage which had ingested an altered lymphocyte or neutrophil leucocyte (110).

In the following year Klemperer (111) called attention to a lesion in the diseased organs which had been observed previously but not much attention had been paid to it. This was the presence of clumps of haematoxylin staining bodies, in the heart valves, in the connective tissue and sometimes in blood vessels which had been interpreted in the past as degenerate cells. Klemperer believed that they were a more specific lesion than the fibrinoid changes and by histochemical technique concluded that these bodies were derived from cell nuclei and consisted of depolymerized desoxyribose nucleic acid and the inclusion body in the L.E. cell was of a similar nature, he then went further and showed by morphological and histochemical studies that there was a relationship between the fibrinoid

breaks in the elastica and disappearance of mast cells. Bellingham, Orr and Woodhouse (101) have shown by transplantation experiments that it is the stroma which is the determining factor in certain types of experimental carcinomata. The next stage in malignancy, that of infiltration, has been studied by Gersh and Catchpole (102) and Sylven (103). They have found that in a rapidly-growing tumour there is marked depolymerization of the ground substance with increased serum glycoproteins, and the reverse is to be seen in slow-growing tumours where there is evidence of a fibrous reaction. The contrast between a scirrhous tissue reaction and the lack of tissue reaction in an encephaloid carcinoma is familiar to all, but the difficulty is that given these morphological observations, one is uncertain how to interpret them; is it *post hoc* or *propter hoc*? The answer may partially be resolved by the introduction of more critical methods for detecting depolymerization and proteolytic enzymes.

Disseminated Lupus Erythematosus

Although there is much more that I would like to tell you of the pathology of connective tissue—the significance of changes in atherosclerosis and aortic medial necrosis where one has an admirable example of depolymerization taking place in relation to all the elements of connective tissue (104), and the nature and possible mechanism of the changes in the ill-named collagen diseases (105), in conclusion I would like to say a few words about disseminate lupus erythematosus, not so much from the point of view of its special pathology, but because it is an example of a disorder characterized by lesions in connective tissue in which we are beginning to get some idea of pathogenesis.

Disseminate lupus is a rather uncommon disease (106), chiefly seen in adult women. The patient presents a variety of vague symptoms of which the most constant is irregular fever with malaise and weakness, there will be attacks of swelling of the joints and very frequently a rash, particularly on the malar region of the face which is red and scaly. I shall not describe the clinical signs or laboratory findings, but ordinarily the disease runs an irregular course with exacerbations and remissions over months or years, and until recently was invariably fatal. At

3. KLEMPERER, P. The Concept of Collagen Disease. *Amer J. Path.* **26**, 505. (1950)
4 RANDALL, J. T., FRASER, R. D. B, JACKSON, S., MARTIN, A. V. W. and NORTH, A. C. T. Aspects of Collagen Structure. *Nature*, **169**, 1029 (1952).
4. BEAR, R. S. The Structure of Collagen Fibrils. *Advances in Protein Chemistry*, **7**, 69. (1952).
5. KRAMER, H. and LITTLE, K. Nature of Reticulin. *Nature*, **170**, 499 (1952).
6. WOLPERS, C. Zur Elektronen mikroskopischen Darstellung Elasticher Gewebselemente. *Klin. Woch.* **23**, 169 (1944).
7. GROSS, J. The Structure of Elastic Tissue as studied with the Electron Microscope. *J. exp. Med.* **89**, 699. (1949).
8. GROSS, J. Fibre formation in trypsinogen solutions: An Electron Optical Study. *Proc. Soc. exp. Biol. Med.* **78**, 241. (1951).
8. FRANCHI, C. M. and DE ROBERTIS, E. Electron microscopic observations on Elastic Fibres. *Proc Soc. exp. Biol. Med.* **76**, 515. (1951).
9. HALL, D. A., REED, E. and TUNBRIDGE, R. E. Structure of Elastic Tissue *Nature*, **170**, 264. (1952).
10. LANSING, A. I., ROSENTHAL, T. B., ALEX, M. and DEMPSEY, G. W. The Structure and Chemical Characterisation of Elastic Fibres as revealed by elastase and Electron Microscopy. *Anat. Rec.* **114**, 555. (1952).
11. TUNBRIDGE, R E, TATTERSALL, R. N., HALL, D. A, ASTBURY, W. T. and REED, R. The Fibrous Structure of the normal and abnormal Human Skin *Clin Sci.* **11**, 314. (1952).
12. COWDREY, E. V. *Principles of Ageing* Chapter 23 'Ageing of Tissue Fluids'. p. 686 London, 1939
13. BENSLEY, S. H. On the presence, properties and distribution of the intra-cellular ground substance of loose Connective Tissue. *Anat Rec.* **60**, 93. (1934).
14. PEARSE, A G E *Histochemistry, theoretical and applied* p. 134 London, 1953.
15. LOEWI, G Changes in the ground substance of ageing cartilage. *J. Path Bact.* **65**, 381 (1953).
16. EDLUND, T Studies on Absorption of Colloids and fluid from rabbit knee joints *Acta physiol scand* **18** Suppl 62, Uppsala, 1949.
17 MCMASTER, P D An Enquiry into the structural conditions affecting fluid transport in the interstitial tissue of the skin *J. exp Med.* **74**, 9 (1941)
18 GUTMAN A B. and YU, T. F A Concept of the role of enzymes in endochondral calcification. *Trans 2nd Macy Conf on Metabolic Interrelations* (1950) 167. New York, 1950
19 HENDRELLS, S. B and HILL, W. L. The nature of bone and phosphate rock *Trans. 3rd Macy Conf on Metabolic Inter-relations* (1951). 173. New York, 1951

material and the haematoxylin bodies and that the fibrinoid consisted of a deposition in connective tissue of histones and other protein residues derived from extreme depolymerization of nuclear proteins, though this was not to say that fibrinoid occurring in other conditions was of the same type (112). Thus it appeared that there was present in the serum of patients with this disease a substance which was able to induce depolymerization of nucleoproteins; then last year it was shown that the 'L.E. factor' was an inactivator of the desoxyribose nuclease inhibitor normally present in cells (113), and that the factor reacts with the inhibitor and allows nucleoprotein depolymerization by the desoxyribose nuclease which is in the cytoplasm and ordinarily plays a part in the controlled energy metabolism of the cell (114). So there is evidence, not all proved, of an intrinsic biochemical lesion in this obscure disease and it is easy to understand that widespread nucleoprotein breakdown will have dire consequences on the patient's metabolism and general well being. What is the origin of the L.E. factor we do not know but this example serves to illustrate that a cellular disturbance can induce changes in connective tissue and is a reaffirmation of Virchow's dictum 'omnis cellula e cellula'.

In the controlled release of atomic energy, it is necessary to have a substance, known as a moderator, which slows down and to some extent controls the fast neutrons so that they are captured by the fissionable material, and allows the chain reaction to proceed. Korsshelt has described connective tissue as a temporal document on which the life history of the organism is indelibly inscribed; it is much more than that, it is the biological moderator of cellular energy.

REFERENCES

1. ASTBURY, W. T. Adventures in Molecular Biology. Harvey Lectures, Series 46, 3. (1952).
2. KLEMPERER, P., POLLAK, A. D. and BAEHR, G. Diffuse Collagen Disease. *J.A.M.A.* 119, 331 (1942).

38. HASS, G. M. Elastic Tissue. *Arch. Path.* **27**, 334; 583. (1939).
39. BOURNE, G. H., Editor. *Cytology and Cell Physiology.* 2nd edition. Oxford, 1951.
40 HASS, G. M. Membrane Formation at Lipoid-aqueous interfaces in tissues. *Arch, Path.* **27**, 15. (1939).
41. MONTGOMERY, G. L. Healing of experimental wounds of lung *Brit. J. Surg.* **31**, 292. (1943/4).
42. GERSH, L. and CATCHPOLE, H. R. The organisation of ground substance and basement membrane and its significance in tissue injury, disease and growth. *Am. J. Anat.* **85**, 457. (1949).
43. ASBOE-HANSEN, G. The original of synovial mucin. *Ann. Rheum. Dis.* **9**, 149. (1950).
44 RILEY, J. F. The effects of Histamine liberators on the mast cells of the rat. *J. Path. Bact.* **65**, 471. (1953).
45. ELSTER, S. K. Effect of Ascorbic Acid deficiency on collagen content
46.
47. ticular cartilage related to function. *Brit. med. J.* **2**, 1295 (1952).
48 KING. L. S. Vital staining of connective tissue. *J. exp Med.* **68**, 63. (1938).
49. BARCROFT, J., DANIELLI, J. F., HARPER, W. F. and MITCHELL, P. D. Wharton's Jelly considered as a conducting path *Nature*, **154**, 667. (1944).
50. DAY, T. D The mode of reaction of interstitial connective tissue with water *J. Phys.* **109**, 380. (1949).
51. MCMASTER, P. D and PARSONS, R. J. The movement of substance and the state of the fluid in the intradermal tissue *Ann. N.Y. Acad. Sc.* **52**, 992. (1950)
52. OPIE E. L. and ROTHBARD, M. B. Water exchange of collagenous tissues and gelatin. *J. exp. Med.* **97**, 409. (1953).
53. ROLLHAUSER, H. Experimentale Beeinflussing der Dehnbarkheit des Bindegewebes am lebenden organismus. *Klin. Woch.* **26**, 126. (1948).
54. ALTSHULER, C. H and ANGEVINE, D. M. Acid mucopolysaccharides in degenerative disease of connective tissue with special reference to serous inflammation. *Am J. Path.* **27**, 141 (1951).
55 BLOOM, W, BLOOM, M A. and MCLEAN, F. C. Calcification and ossification Medullary bone changes in the reproductive cycle of female pigeons *Anat Rec.* **81**, 443 (1941).
56 ENGEL, M B Mobilisation of mucoprotein by parathyroid extract.
57. 1096 (1952).
58. ASBOE-HANSEN, G The intercellular substance of the connective tissue in myxoedema *J investig Dermat* **15**, 25. (1950).

20. ROBINSON, R. A. and WATSON, M. L. Collagen crystal relationship in bone as seen in the electron microscope. *Anat Rec.* **114**, 383. (1952).
20. RUTISHAUSER, E. and MAJNO, G. Physiopathology of Bone Tissue. *Bull. Hosp. Joint Dis.* **12**, 468. (1951).
21. BAITSELL, G. A study of the development of connective tissue in amphibia. *Am. J. Anat.* **28**, 447. (1921).
22. STEARNS, M. L. Studies on the development of Connective Tissue in transparent chambers in the rabbit's ear. *Am. J. Anat.* **66**, 133 (1940). **67**, 55. (1940).
23. DOLJANSKI, L. and ROULET, F. Studien uber die Enstehung der Bindegewbsfibrille. *Virchow's Arch. f. Path. Anat.* **291**, 260. (1933).
24. PORTER, K. R. Repair processes in Connective Tissue. *Trans. 2nd Macy Conf. on Connective Tissue* (1951). 126. New York, 1952.
25. WYCKOFF, R. W. G. The fine structure of Connective Tissue. *Trans. 3rd Macy Conf on Connective Tissue* (1952). 38. New York, 1952.
26. OREKHOVICH, V. N., TUSTANOVSKII, A. A., OREKHOVICH, K. D. and PLOTNIKOVA, N. E. Isolation of crystalline proteins of a new type (procollagen) from various organs of vertebrates. *Biochimia*, **13**, 55 (1948)
27. HIGHBERGER, J. H., GROSS, J. and SCHMITT, F. O. Interaction of mucoprotein with soluble collagen, an electron microscope study. *Proc. Nat. Acad. Sc. Wash* **37**, 286. (1951).
28. SCHMITT, F. O., GROSS, J. and HIGHBERGER, J. H. A new particle type in certain connective tissue extracts *Proc. Nat. Acad. Sc. Wash* **39**, 459 (1953).
29. HARKNESS, R. D., quoted by NEUBERGER, A. Protein Metabolism. *Brit med Bull.* **8**, 210. (1952).
30. MANCINI, R. E. and DE LUSTIG, E. S. Accion de la Hialuronidasa sobre los mucopolisacarados de los fibroblastos 'in vitro'. *Rev. Soc. argent. Biol* **26**, 227 (1950).
31. FISCHER, A. The participation of Blood Proteins in the metabolism of Tissue Cells *Biol Reviews*, **22**, 178 (1947).
32. HAUROWITZ, F. and CRAMPTON, C. F. The role of the nucleus in protein synthesis *Exper. Cell Res* Suppl. **2**, 45 (1952).
33. PENNEY, J. R. and BALFOUR, B. M. The effect of Vitamin C on mucopolysaccharide production in wound healing *J. Path Bact* **61**, 171. (1949).
34. PIRANI, C. L. and CATCHPOLE, H. R. Serum glycoproteins in experimental scurvy *Arch. Path.* **51**, 597 (1951)
35. BRADFIELD, J. R. S. Phosphatase Cytochemistry in relation to protein structure. *Exper Cell. Res* Suppl **1**, 338 (1949).
36. KRAMER, H. The nature and formation of normal and abnormal connective tissue D. Phil thesis, University of Oxford (1952).
37. MERCER, E. H. The fine structure and biosynthesis of silk fibroin *Austr J. Sc. Res.* Series B, **5**, 366. (1952)

73. WOLBACH, G. B. and MADDOCK, C. L. Cortisone and Matrix formation in experimental scorbutus, and repair therefrom. *Arch. Path.* **53**, 54. (1952).
74. ALLBRIGHT, F. Cushing's Syndrome. Harvey Lectures, Series **38**, 123, (1943).
75. HENCH, P. S , KENDALL, E. C , SLOCUMB, C. H. and POLLEQ, H. F. The effect of a hormone of the adrenal cortex (17-hydroxy-11 dehydrocorticosterone, Compound E) and of pituitary adrenocorticotropic hormone on rheumatoid arthritis; preliminary report. *Proc. Staff Meet. Mayo Clin.* **24**, 181. (1949).
76. ROBERTS, E , KARNOFSKY, D. A. and FRANKEL, S. Influence of corti-
77. L[illegible]
animal tissues. *Proc. Soc. exper. Biol. Med.* **76**, 596. (1951).
78. ASBOE-HANSEN, G The mast cell an object of cortisone action on Connective Tissue. *Proc. Soc. exp Biol. and Med.* **80**, 677. (1952).
79. BLOOM, F. Effect of cortisone on Mast cell tumours (Mastocytoma) of the dog. *Proc. Soc. exp. Biol. and Med.* **79**, 651. (1952).
80. TAUBINHAUS, M and AMROMIN, G. D. Influence of steroid hormones on granulation tissue. *Endocrinology*, **44**, 359. (1949).
81. HERINGA, S C , LEYNS, W. F. and WERDINGER, A. Water adsorption of the connective tissue of cornea in dependence on adrenal insufficiency. *Acta brev. neerl. Physiol.* **10**, 49 (1940)
82. KULONEN, E. On the relation of hyaluronic acid to the water and electrolyte metabolism. *Acta Physiol. Scandinav.* **27**, 82. (1953).
83. WELLS, H G. *The Chemistry of Arteriosclerosis.* Cowdry's Arteriosclerosis. New York, 1932.
84 WEISS, P. *Differential Growth* Chapter 7. 'The Chemistry and Physiology of Growth' Princeton, 1949.
85. SPRUNT, D H The Ground Substance in Infection *Ann. N.Y Acad. Sc* **52**, 1052. (1950).
86 MEYER, K. and RAPPORT, M. M. Hyaluronidases. *Advances in Enzymology*, **13**, 199. New York, 1952
87. HAAS, E On the mechanism of Invasion *J biol. Chem* **163**, 63, 89, 101. (1946).
88. BURNET, F M. Biological properties of the mucoprotein urinary inhibitor. *Austr. J exp Biol. and Med.* **30**, 251. (1952)
89. GITLIN, D , LANDING, B. H. and WHIPPLE, A M The localisation of homologous plasma proteins in the tissues of young human beings as demonstrated with fluorescent antibodies *J exp Med* **97**, 163 (1953)
90 ALTSHULER, C. H and ANGEVINE, D M. Histochemical Studies on the pathogenesis of fibrinoid. *Amer J. Path.* **25**, 1061 (1949)
90. GLYNN, L E and LOEWE, G. Fibrinoid necrosis in rheumatic fever. *J Path Bact* **64**, 329 (1952)

59. PALITZ, L. L. and BRUNNER, M. J. The Mucinoses: a classification with histochemical studies on the nature of mucin. *J. investig. Dermat.* 14, 159. (1950).

60. ASBOE-HANSEN, G. and WERSEN, K. Influence of thyrotrophic hormone on connective tissue. Pathogenetic significance of mucopolysaccharides in experimental exophthalmos. *Acta Endocrin.* 8, 90. (1951).

60. ASBOE-HANSEN, G., WERSEN, K. and WICKMANN, R. Malignant exophthalmos muscular changes and thyrotropin content of serum *Acta Endocrin.* 11, 376. (1952).

60. LUDWIG, A W, BOAS, N. F. and SOFTER, L. J. Role of mucopolysaccharides in pathogenesis of malignant exophthalmos. *Proc. Soc. exp. Biol. Med.* 73, 137. (1950).

61. LOEB, L, SUNTZEFF, J. and BURNS, E. L. Changes in the nature of the stroma in vagina, cervix and uterus of the mouse produced by long continued injections of estrogen and advancing age. *Am. J. Cancer,* 35, 159 (1939).

62. HARKNESS, M. L. R. and HARKNESS, R. D. Collagen in the reproductive tract of the rat during pregnancy and lactation. *J. Phys.* 120, 7 P. (1953).

63. LIPSCHUTZ, A. *Steroid hormones and tumours.* Baltimore, 1950.

64. ELLIOTT, R. I K. and ROBB-SMITH, A. H. T. Unpublished observations.

65. TALMADGE, R. V. A histological study of the effects of relaxin on the symphysis pubis of the guinea pig *J exp. Zool* 106, 281. (1947).

66 PETRY, G. Die Konstruktion des Eierstockbindegewebes und dessen Bedentung fur den Ovariellen Zyklus. *Zeit. f. Zellforsch.* 35, 1. (1950)

67. DURAN-REYNALS, F, BUNTING, H, VAN WAGENIN, S. Studies on the sex skin of macaca mulata *Ann. N Y. Acad. Sc.* 52, 1006 (1950).

68. HARDESTY, M. The structural basis of the response of the comb of the brown Leghorn fowl to sex hormones *Amer. J Anat.* 47, 277. (1931).

69. DURAN-REYNALS, F Introduction to Conference on the ground substance of the mesenchyme and hyaluronidase. *Ann N.Y. Acad Sci.* 52, 946 (1950).

70. CATCHPOLE, H R Solubility properties of some components of the ground substance in relation to intravital staining of connective tissues. *Ann N.Y. Acad. Sci.* 52, 989. (1950)

71. RAGAN, C Effect of ACTH and cortisone on Connective Tissue. *Trans 1st Macy Conf on Connective Tissue (1950).* 137 New York, 1951

72. CAMERON, G. R A survey of tissue responses to ACTH and Cortisone. *The Suprarenal Cortex*, Edited by J M. Yoffey London, 1953.

72. BANGHAM, A D. The effect of Cortisone on Wound healing. *Brit. J. exp. Path* 32, 77. (1951).

107. KLEMPERER, P., POLLACK, A. D. and BAEHR, G. Pathology of Disseminated Lupus Erythematosus. *Arch. Path.* 32, 569. (1941).

108. HARGREAVES, M. M. and RICHMOND HAND MORTON, R. Presentation of two bone marrow elements: The 'Tart' cell and the 'L.E.' cell. *Proc. Staff. Meet. Mayo Clinic*, 23, 25. (1948).

109 HASERICK, J. R. and LEWIS, L. A. Induction of specific antibodies against L.E. factor. *Blood*, 5. (1950).

109. BERMAN, L., AXELROD, A. R., GOODMAN, H. L. and MCCLAUGHRY, R. I. So-called 'lupus erythematosus inclusion phenomenon' of Bone, Marrow and Blood. *Amer. J. Clin. Path.* 20, 403. (1950).

109. SMITH, P. A. J. The 'L.E. cell' and its significance. *Brit. J. Derm.* 64, 10. (1952).

110. ROHN, R. J. and BOND, W. H. Some supravital observations on the 'L E.' phenomenon. *Amer. J. Med.* 4, 422. (1952).

111. KLEMPERER, P, GUEFT, B. and LEE, S. Nucleic acid depolymerization in systemic lupus erythematosus. *J. Mt Sinai Hosp.* 16, 61. (1949).

111. KLEMPERER, P., GUEFT, B., LEE, S., LEUCHTENBERGER, C. and POLLISTER, A. W. Cytochemical changes of acute lupus erythematosus. *Arch. Path.* 49, 503 (1950).

112. KLEMPERER, P. Pathology of systemic lupus erythematosus. Chap. 3 in *Progress in Fundamental Medicine*. Ed. by J. F. A. McManus. London, 1952.

113. HENSTELL, H. H. and FREEDMAN, R. I. An inhibitor of desoxyribonuclease in human white blood cells and bone marrow cells and its relationship to cellular maturity. *Science*, 115, 357. (1952).

114. KURNICK, N. B., PARISER, S, SCHWARTZ, L. I, LEE, S. L. and IRVINE, W. Studies of desoxyribonuclease in systemic lupus erythematosus *J clin. Invest.* 31, 1037 (1952).

114 KURNICK, N B. SCHWARTZ, L., PARISER, S, LEE, S and IRVINE, W. The role of desoxyribonuclease and a nuclease inhibitor from leucocytes in the lupus erythematosus cell phenomenon. *J. clin. Invest* 31, 645. (1952).

91. STUART, E. G. The Mast Cell in Anaphylaxis. *Anat. Rec.* **112**, 394. (1952).
92. BENSLEY, S. H. The origin of the mast cell. *Anat. Rec.* **112**, 310. (1952).
93. EPPINGER, H. *Die Permeabilitäts-pathologie*. Vienna, 1949.
94. ROBB-SMITH, A. H. T. Discussion on the Collagen Vascular Diseases. *Proc. Roy. Soc. Med.* **45**, 811. (1952).
95. HILL, K. R., RHODES, K., STAFFORD, J. L. and AUB, R. Serous Hepatosis: a pathogenesis of hepatic fibrosis in Jamaican Children. *Brit. med. J.* **1**, 117. (1953).
96. TEILUM, G. Cortisone-Ascorbic Acid interaction and the pathogenesis of amyloidosis. *Ann. Rheum. Dis.* **11**, 119. (1952).
97. COMAN, D. R., MCCUTCHEON, M. and ZEIDMAN, I. Failure of Hyaluronidase to increase the invasiveness of a neoplasm. *Canc. Research*, **7**, 383. (1947).
98. SIMPSON, W. L. Mucolytic enzymes and invasion by carcinomas. *Ann N.Y. Acad. Sc.* **52**, 1125, (1950).
99. CRAMER, W. and SIMPSON, W. L. Mast cells in experimental skin. Carcinogenesis. *Canc. Research*, **4**, 601. (1944).
100. ORR, J. W. The changes antecedent to tumour formation during the treatment of mouse skin with carcinogenic hydrocarbons. *J. Path.*

101.

methyl-cholanthrene. *Brit J Canc.* **5**, 417. (1951).
102. GERSH, I. Ground substance and the plasticity of connective tissue. Harvey Lectures, Series **45**, 211. (1952).
103. SYLVEN, B. Changes in cartilage produced by infiltrating carcinoma. *Brit. J Canc* **1**, 103 (1947)
103. SYLVEN, B. Ester sulphuric acids in stroma connective tissue. *Acta Radiologica*, **32**, 11 (1949).
104. MOON, H. D and RINEHART, J. F. Histogenesis of coronary arteriosclerosis *Circulation*, **6**, 481. (1952).
104. BUNTING, C. H and BUNTING, H Acid mucopolysaccharides of the Aorta. *Arch. Path.* **55**, 257. (1953).
104. TAYLOR, H G A study of the Human Aorta dealing with the role of mucopolysaccharides in the pathogenesis of internal fibrosis and atherosclerosis. *Amer. J Path.* **29**, 589. (1953).
105 EHRICH, W. E Nature of Collagen Diseases *Am. Heart J.* **43**, 121. (1952).
105 DUFF, G. L. The Collagen Diseases The Musculo-Skeletal System. Ed. M. Ashford New York, 1952.
106. CLUXTON, H. E and KRAUSE, L. A M. Acute lupus erythematosus disseminatus *Ann Int Med* **19**, 843. (1943).
106. COHEN, H. and CADMAN, E F B. The natural history of Lupus Erythematosus. *Lancet*, **2**, 305. (1953).

are found to occupational diseases including those due to the inhalation of dust. Georgius Agricola (1494–1555) in his treatise on mining and metallurgy *De Re Metallica* mentioned the dangers and diseases to which miners and refiners were prone. Paracelsus, a Swiss alchemist and physician (1493–1541), wrote concerning the diseases of miners and smelter-workers (1567). Dimerbrock (1649) was perhaps the first to cut sections of a stone-cutter's lung, and he described how the organ was clogged with fine stone dust. Ramazzini, known as the third Hippocrates and the father of industrial medicine, described in his *De Morbus Artificium Diatriba* (Padua, 1700) many occupations and the special hazards involved in each, including dust diseases and their pulmonary lesions.

It is an interesting speculation as to whether prehistoric man engaged in the breaking and shaping of flints for arrowheads and spearheads may have suffered from silicosis. In many parts of England, particularly in the Eastern counties, there are extensive deposits of man-made flints fashioned in the shape of weapons and tools, whose production must have involved the expenditure of considerable energy and the production of much finely-divided, free silica dust. It is certain, however, that in later years the flint-knappers, who were extensively employed in the musket trade, inhaled large quantities of finely-divided flint dust and suffered from silicosis. But it was with the introduction of industrial processes, handling large amounts of flint, that the disease became a serious social problem Thomas Benson, of Newcastle-under-Lyme, was granted patents for the wet grinding of flint in 1726 and 1732 because it was found that when powdered dry the process proved 'very destructive of mankind. Any person ever so healthful or strong working in that business cannot probably survive above two years, occasioned by the dust thrust into his body by the air he breathes, which, being of a ponderous nature, fixes there so closely that it is now very difficult to find persons which will engage in the business, to the great obstruction and detriment of the said trade.' This powdered flint was used as a grinding powder and for the manufacture of grindstones. A Sheffield practitioner (Knight, 1830) stated that '*Grinders who use a dry grindstone*

VII

Silicosis

E. J. KING

PNEUMOCONIOSIS (from the Greek) is a 'dusty condition of the lungs', and implies both the inhalation of excessive amounts of dust and the deposition of large quantities of fine dust particles in the pulmonary tissue. Silicosis is a pneumoconiosis due to inhalation of finely powdered siliceous rock, sand or flint. Workers who breathe dusty atmospheres containing fine particles of silica rock suffer from cough, rigidity of the chest, shortness of breath, dyspnoea and emphysema.

Industries in which silicosis has been common are flint breaking, mining, quarrying, grinding with silica abrasives, potteries and refractories, foundries, and so on. The lungs of such workers become loaded with stone dust and the normal alveolar tissue is largely replaced by fibrous scar tissue. This scar tissue assumes the characteristic form of silicotic nodules illustrated in Plate II, Fig. 1,[1] which is a section through the lung of a stone-worker.

HISTORY OF SILICOSIS

The history of silicosis is of very ancient origin. Its existence was recognized in the first descriptions we have of occupational diseases made before the Christian era. Hippocrates (460–370 B.C.) described symptoms in metal-diggers which compare well with more modern accounts of pneumoconiosis amongst miners. Pliny's (A.D. 23–79) description of a protective mask for miners contains another early reference to the harmfulness of dust in the lungs. In Galen's (A.D. 131–201) writings many references

[1] The plates referred to in this lecture will be found at the end of the book.

which means 'lung full of dust'. He described in detail the anatomy of dusty lungs; and Kussmaul gave the chemical proof of the siliceous nature of the dusts in the lungs. Visconte (1870) in Italy was the first to use the term 'silicosis' to denote the pathological changes in the lung due to the inhalation of siliceous dust.

Detailed accounts of the progress of dust disease amongst stone-cutters are to be found in the writings of Alison and Hugh Miller (1869). They observed that stone-drillers appeared to suffer more from the rapid onset of the disease than did other miners, and the conviction gradually arose that it was only in workers with hard stone that dust disease seriously affected health. Workers in softer stone seemed to suffer less severely from lung diseases, and the pathological changes produced were not so gross (Plate III, Fig. 2 and Plate IV, Fig. 3). Thus workers with shale and clays of various sorts and in haematite mines were not so frequently or so early in life incapacitated from work; and the fibrosis in their lungs was less nodular and more diffuse than that observed in the lungs of flint-, quartz- and granite-workers. These observations led to the theory that not all dusts of different kinds of stone were equally pathogenic. Early in the present century Haldane recognized that there must be a difference between siliceous mineral dusts. Microscopically the particles of all of them appeared to be hard and sharp, but they were not equally destructive of health. Some stone dusts even appeared to be antidotal to others.

Theories of Silicosis

Mechanical theory. Early workers in the field thought that dust was harmful because of its physical presence in the lungs; that the sharpness and hardness of its particles produced a sort of microscopic wounding or trauma of the delicate tissues. Haldane's observation that the practice of stone-dusting in coal mines to allay explosions had not increased the incidence of lung disease, but appeared rather to lessen it, led to some doubt as to whether the physical presence of the dust itself could produce the pathological effect observed It might be that some dusts damaged the lungs because they contained a toxic substance which was absent from other dusts.

die at the age of 28–32 years'; and suggested that this sort of work should be reserved for convicted criminals.

In 1832 Thackrah studied the expectation of life of British workers in various dusty trades. His most valuable conclusion was that not all dust shortened life; workers in sandstone quarries generally died before they reached forty years of age, though there was hardly any lung disease among brick and lime stone-workers. Thackrah was a great advocate of prevention. He advised greater ventilation, face masks and wet processes of working. He anticipated the modern practice of infusing the working face in coal-mines with water: 'Particles are laid as they are formed, by the continuous oozing, dropping and splashing of the insinuated water.' So that the medical man might not be accused of interfering in matters which did not concern him, he remarked, 'If any object that the cure, not the cause or prevention of disease, is the business of the medical practitioner, I would reply that the scientific treatment of a malady requires a knowledge of its nature, and the nature is but imperfectly understood without a knowledge of the cause'.

The pathology of coal-miner's pneumoconiosis was described by Gregory (1831) and Marshall (1834); and Thomson drew attention to the breathlessness, asthma and the black sputum common to workers in coal-mines. In 1837 he accurately described the pathological appearance of the hard, black lungs of coal-miners, and was aware of the condition associated with working in stone dust Peacock (1861) was probably the first physician in Great Britain to distinguish between miner's phthisis and pulmonary tuberculosis; the view of current medical opinion being that miner's phthisis was caused by bad hygiene rather than by dust In 1865 Greenhow studied the pathological changes in the lungs he had collected from stone-workers and described the gritty feeling of the lungs when they were cut with a knife, and referred to the appearance of microscopic drifts and deposits of sand in the lung tissue. From the lungs of one worker he isolated about 30 g. of this so-called sand; and he was perhaps the first to use polarized light to demonstrate fine silica particles in the lungs of metal-grinders. Zenker (1866) first introduced the term 'pneumonokoniosis'

quartz and other forms of silica such as flint would release silicic acid ('soluble silica') into solution; and the modern 'solubility theory of silicosis' supposes that it is this substance which irritates the tissues of the lungs. The response of the lung to the constant release of silicic acid from the slowly dissolving particles is the development of fibrous tissue, and for this purpose close contact of the particles with the tissue is important.

Mavrogordato (1926) and Policard (1933) pointed out that these slowly dissolving particles had a most peculiar effect on living tissue. Cells poisoned by them did not tend to disintegrate and disappear as do cells killed by other means; they became 'mummified' and preserved in some way which may be roughly analogous to the preservation of eggs in waterglass.

Solubility of Silica

In his well-known monograph on silica and the silicates, Sosman (1927) has said that from the standpoint of the industrial chemist silica is about the last thing on earth which would be expected to be harmful to the body Its inertness is a matter of common knowledge, and in its crystalline forms such as quartz it is usually thought of as being one of the least soluble of substances This current belief in the insolubility of quartz does not, however, stand the test of experiment. It is true that if a large quartz pebble is kept in a vessel of water no silica can be demonstrated to have passed into solution, but if the quartz be reduced to a very fine powder and then suspended in water the presence of significant amounts of silica in solution can be demonstrated.

Many of the older results recorded for the solubility of quartz in water are of little value because they were obtained at a time when there were no adequate means of separating the most finely divided of the quartz particles from the solution. The filtrates on which the measurements of solubility were made contained suspended particulate matter in addition to the dissolved silica Many early results were therefore unreliable and did not provide a proper measure of the dissolved silica. Modern methods, however, have removed these difficulties. By means of the ultracentrifuge and the ultrafilter it is possible to obtain true solutions completely separated from the suspended solid matter (cf.

Chemical theory. This reasoning led Gye and Purdy (1922) and Gye and Kettle (1922) to advance their well-known chemical theory of silicosis which proposed that silica dust was harmful in the lungs not because it was hard and sharp, but because it was toxic; and produced a soluble poisonous substance which irritated the surrounding tissue and cells in which the particles were engulfed. It was because the components, or some of the components, of this dust were sufficiently soluble to release some of this soluble silica (silicic acid) into the fluids of the engulfing cells that these dusts were harmful; and other dusts were harmless or less harmful because they did not contain components which were capable of releasing as much, or any, silica into solution.

Gardner's experiments (1923) with carborundum showed it to be non-silicosis-producing. Its particles were exceedingly hard, sharp and abrasive, and yet they did not produce silicotic nodules in the lungs of human beings or animals. On the mechanical theory, carborundum should have produced a silicosis-like condition; but only a marked foreign body reaction and diffuse fibrosis of the lungs was observed. Since then industrial experience has confirmed that carborundum is not a disease-producing dust in the sense that silica is. Kettle (1932) showed that the same quartz particles which readily produced silicosis in animals would no longer do so if they were first coated with a thin layer of iron oxide which did not alter their microscopic appearance or sharpness. By the precipitation of ferric chloride with ammonia in the presence of powdered quartz, he succeeded in getting iron hydroxide to precipitate on the surface of the quartz particles. Despite their microscopic similarity to the uncoated particles, they produced very little effect in animals, whereas the uncoated particles would produce massive fibrosis in the lungs of guinea-pigs in the course of a few months. This seemed to indicate that covering quartz particles would prevent them from producing silicosis. Quartz is a form of pure silica. Evidently a deposit of iron on its surface had rendered it innocuous. This confirmed the idea that the harmfulness of silica is related to the release of a toxic substance from the surface of its particles. Kettle postulated that finely ground

degree of solution. Table 1 illustrates the degree of solution which pure powdered quartz of different particle sizes may undergo when it is shaken for long periods with buffered Ringer's solution. The smallest fraction dissolved to the extent of 14·8 mg./100 ml., and in 20 weeks a total of 10 per cent of the original solid matter had passed into solution.

Silica occurs in two forms in nature — free silica where the mineral consists of pure SiO_2, and silicates. Quartz, tridymite, cristobalite, flint and opal are typical forms of free silica. These dissolve to about the same extent to give 13–15 mg. SiO_2 per 100 ml. of pH 7·4–buffered Ringer's solution, but the amount dissolved is very definitely related to the size of the particles (Table 2, King *et al.*, 1953).

TABLE 2. *Silica solubility of flint dusts of different size*

Size of particles ..	<0 5μ	0 5–1μ	1–2μ	2–4μ	4–8μ
mg.SiO_2/100 ml. ..	13 6	13 4	13 1	8 9	5 1

Granite forms of rock are rich in free silica; and many metals which are mined occur in close association with quartz. Other rocks such as shale and slates and much country rock, which do not seem to produce silicosis, consist mainly of silicates with little free silica. The 'silica solubility' of the silicates appears to be almost uniformly much less than that of the free forms of silica when the powdered minerals of similar particle size distribution are tested. Felspar, Fuller's earth, kaolin, mica, olivine, sericite, talc, etc , and many natural mixtures of silicates in various forms of rock, yield only about 1 mg of SiO_2 into solution in 100 ml. Ringer's solution and blood plasma. We shall see later that a rough parallelism exists between the amount of silica which various minerals yield into solution and their apparent pathogenicity when introduced into the lungs of animals.

Histological Characteristics of Human Silicosis

Human silicosis presents a histological picture with very definite characteristics. The small nodular foci of fibrous tissue assume in a well-developed nodule a 'ball-of-twine' appearance in which there is far more fibrous tissue than appears necessary to

King and McGeorge, 1938; Titus, 1937). The utilization of chemical reactions in which only the truly dissolved silica will react has furnished a further means of measuring the actual amount of dissolved silica. The silico-molybdic acid procedure of King (1928, 1939, 1947) and others has made it possible to measure the chemically reactive silica and to gain a measure of the amount of silica present in solution either as silicic acid or as simple polymers of low molecular weight. The amount of silica found in solution by this method, in the filtrate prepared from a suspension of powdered quartz in buffered water and filtered through the best finest-pored filter paper, and in the supernatant of the same suspension after prolonged centrifuging at 3,000 r.p.m., is the same, and none of this soluble silica is lost after prolonged centrifuging at much higher speeds, e.g. 20,000 r.p.m. Ultra- [illegible] under high pressure [illegible] the loss of a small an[illegible] seem to indicate that there is some polymerized silicic acid present, but that it must be of fairly small dimensions and capable of chemical reaction with molybdate.

TABLE 1 Dissolution of quartz of different particle sizes (2% quartz shaken with Ringer's solution (pH 7·4, 37°C.); dissolved SiO_2 determined by silico-molybdate method)

Size of quartz particles (μ)	Initial solubility (mg. SiO_2 per 100 ml)	Final solubility (after 20 weeks' extraction)	Total dissolved (as %)
<1	15	1·5	10
1–2	13	1	5
2–5	7	1	3

The figures for the solubility of finely powdered quartz in water and Ringer's solution buffered at pH 7·4 and in blood plasma and ascitic fluid are of the order of 4–14 mg SiO_2/100 ml and differ with the particle size of the quartz used, with the amount of suspended material, the temperature and the length of time the mixture is shaken. Small amounts of impurities in the mineral dust (e.g. in quartz or flint) likewise influence the

Pure silicate dusts thus far tested have produced only very mild reactions very similar to those obtained with powdered shale (see below). Thus, kaolin was found to give only loose reticulinosis, as is shown in Plate VII, Fig. 9; and likewise felspar (Plate VII, Fig. 10). Hornblende also produced a minimum of pathological response (Plate VIII, Fig. 11).

Sericite or mica, despite its bad reputation (W. R. Jones, 1933), caused very little fibrosis. The mineral particles accumulated in foci of dust cells, but without much reticulin formation or any nodule production. There is seen only a focal accumulation of dust and some thickening of the alveolar walls (Plate VIII, Fig. 12).

Acid-treated sericite, in contrast with sericite which has been exposed only to water and neutral salt solutions, furnishes a dramatic contrast. Plate IX, Fig. 13 gives the picture of several nodules produced by sericite which had been heated on the water bath with dilute hydrochloric acid and washed well with water. Although not so severe as quartz nodules, there has nevertheless been a great increase in the pathogenicity of the sericite due to the acid treatment. The solubility of this sericite was also considerably increased over that of the original powder.

Mixed dusts of free silica and silicates give lesions in the lungs of animals which are always intermediate between those obtained by quartz on the one hand, and silicates on the other. If there be a high proportion of free silica, silicotic nodules will result, as in the case of the high quartz-sandstone referred to above. But if the silicates (mica, kaolin, felspar, etc.) predominate, then the tissue reaction is mild and like that produced by the silicates themselves; and the activity of the contained free silica seems to be suppressed. Such natural mixtures of minerals have been tested in the form of the dusts from several sandstones, granites, slates, shales, horneblende schist, porphyritic schist, etc.

Sandstone of high quartz content (71 per cent quartz, 22 per cent mica, 2 per cent kaolin) produced nodules almost as severe

[illegible] as well as on its distribution

hold the dust content of the lesion. This great excess of reticulin and collagen fibres in the fibrous tissue has been referred to by Belt, Ferris and King (1940) as 'redundant' fibrosis. There is, moreover, a definite pattern of dust distribution within the silicotic nodule. A central core of densely packed mineral particles is surrounded by an almost vacant zone which corresponds to the densest of the whorled collagen fibres; and this again is followed by a halo of dust corresponding to the periphery of the nodule. These points are shown in Plate IV, Fig. 4 and Plate V, Fig. 5.

Experimental Silicosis

It is possible to test the pathogenicity of mineral dusts in animals; and the results so obtained agree on the whole with industrial experience and the reputation which mineral dusts have gained as to whether they produce silicosis or only mild forms of pneumoconiosis.

Free silica dusts. Quartz nodules can be produced in the lungs of animals by the intratracheal technique of Kettle and Hilton (1932) or by the (
Wright, Ray and .
of these 'silicotic'
silicotic nodules and the dust distribution is very similar (Plate V, Fig. 6). But on microscopic examination with high-power magnification certain differences are seen. The haematoxylin-eosin and the silver-stained sections show that the nodule is much less dense and well-organized than the human silicotic lesion. The fibres shown by silver impregnation are less dense and stain black for reticulin rather than golden-brown for collagen, except in the oldest lesions. The pattern of the fibres is, moreover, less well-organized, and the 'ball-of-twine' appearance is not so apparent. In long-standing lesions in animals, however, the appearance of the human silicotic nodule begins to be assumed; and it is true to say that the older the animal silicotic lesion the more closely similar it is to the human (Plate VI, Figs. 7 and 8). (For assessment of fibrosis see footnote [1].)

[1] The basis of the pathological gradings was as follows grade 1, loose reticulin fibrils with no collagen, grade 2, compact reticulin with or without some collagen, grade 3, somewhat cellular, but made up mostly of collagen, grade 4, completely

Naturally-occurring kaolins, such as those represented in shales, contain variable, sometimes large, amounts of free silica, and yet these natural stone dusts have an exceedingly low solubility. The amount of free silica represented in a sample of stone dust would lead to much more solution of silicic acid if suspended alone in the same volume of liquid. This fact has been brought out in work by Whitehouse (1938), by King and Roman (1938) and by Bremner (1939). Admixtures of the stone dust with powdered quartz or flint depress the solubility from 14 mg. of silica per 100 ml. for quartz alone down to about 1 mg. for the mixture.

Whether or not these depressions of solubility can serve as an explanation of the depression of silicosis with mixtures of stone dust and quartz cannot be stated for certain, but it seems very probable, and at any rate they form a suggestive and interesting parallel.

The material in the shale dust which depresses the silica solubility of quartz is, in fact, a small amount of aluminium or aluminium hydroxide which goes into solution in the water to a sufficient extent to deposit on the quartz particles, producing a thin layer of aluminium hydroxide over their surface. Small amounts of pure aluminium will similarly prevent finely powdered quartz from going into solution. The powdered quartz and flint which displayed silica solubilities of 14 or 15 mg. SiO_2/100 ml. now yield less than 1 mg. SiO_2 in solution Many preparations of aluminium, aluminium oxide and aluminium hydroxide have been tried, and they are nearly all effective in this respect.

The 'prevention' of silicosis. A preliminary account by Denny, Robson and Irwin in 1937 of attempts to prevent experimental silicosis with aluminium in rabbits aroused great interest. A later communication in 1939 confirmed the results previously reported. The admixture of very small quantities of aluminium with quartz dust protected the lungs of rabbits from the fibrosis which developed in five months in those exposed to quartz alone In animals with a concentration of metallic aluminium in the lung of 1 per cent no evidence of fibrosis was detected up to a period of 17 months, whereas control rabbits showed

as those obtained with pure quartz. Sandstone of low quartz content (31 per cent quartz, 51 per cent mica, 6 per cent kaolin) produced a marked contrast. The smaller amount of quartz and the larger amount of silicate resulted in a very much less severe lesion with only a very mild fibrous response and few reticulin fibrils, despite the presence of a very high concentration of mineral dust (Plate X, Fig. 14).

The shale dust (Plate X, Fig. 15) showed the same mild fibrous reaction and the same paucity of reticulin fibrils, although the concentration of dust within the lesion was exceedingly high. This shale dust consisted of 8 per cent quartz, 77 per cent mica and 9 per cent kaolin.

Antidotal stone dust. Stone-dusting in coal-mines has been used for a long time in Great Britain to diminish the hazard of explosion. At first it was feared that the introduction of this extra siliceous material into the air of the mines would increase the risk of silicosis, but industrial experience during many years led to the belief that stone dusting, instead of increasing the prevalence of the disease, actually diminished it. On the strength of this experience, Haldane (1917, 1931) advanced the theory that certain mineral silicates offset the action of quartz in its relation to silicosis. Experimental work by Gardner (1938) on the effect of mineral dusts on the lungs and other tissues of animals appeared to indicate that many silicates modify and even retard the action of free silica. Moreover, Belt and King (1945) found that with natural shales and sandstones, the 10–30 per cent of contained quartz, which would certainly produce silicosis if put into the lungs of animals by itself, failed to do so when in admixture with the kaolin and sericite which were present (see above).

Depression of silica solubility. As stated above most of the mineral silicates are much less soluble in body fluids than are the free forms of silica. Kaolin and mica are particularly insoluble. In a mixture of kaolin, mica and quartz powder suspended in a liquid it might be expected that the amount of silicic acid dissolved would be conditioned by the most active member of the mixture, and that in the presence of an excess of the mixed powder the solubility would approach that of quartz taken alone.

dust clouds (Wright, 1950), have been successful in preventing or retarding the production of silicosis in the lungs of rats given aluminium and quartz to breathe, as compared with a like group which rapidly developed silicosis in the lungs from breathing quartz dust without the admixed aluminium.

Silica solubilities of the mixture of quartz and aluminium are given in Table 3, where it will be seen that the dissolution of the quartz powder in the Ringer's fluid was depressed almost to the limit of chemical determination by the presence of aluminium. The quartz failed to stain with aurine when treated with the dye in ammonium acetate buffer, whereas the aluminium-treated quartz stained a bright red which could not be removed by repeated washing, thus demonstrating the presence of a coating of aluminium hydroxide on the surface of the particles.

TABLE 3. Silica solubilities of quartz and quartz + aluminium

	Dissolved silica (mg SiO_2/100 ml)
Quartz + 2% aluminium powder	0 1
„ + 0 5% „ „	0 1
„ + 0 1% „ „	0 2
„ + 0 02% „ „	1 7
„ + 0% „ „	14 0

Animals were killed at regular intervals in order to follow the progress of the changes in the lungs produced by the inhaled dust. Chemical analysis revealed that there was, on the average, approximately the same concentration of quartz introduced into the lungs of the animals in each group. Analysis for aluminium confirmed its presence in the lungs of the quartz + aluminium-dusted animals and its absence from the quartz group.

In the quartz-dusted rats there was no focal reticulinosis found in the lungs during the first 180 days. There was some thickening of a few alveolar walls and some doubtful increase of the perivascular fibrous tissue. At 220 days there were a few focal collections of dust-laden phagocytes which were traversed by a delicate network of fine reticulin. In the hilar glands, however, foci of collagenous fibrosis were found. At 300 days, and thereafter, typical experimental silicotic nodules occurred.

well-developed fibrosis in six or seven months. The inactivation of quartz takes place only when aluminium is closely associated with it in the tissues; the thin impermeable layer of alumina on the quartz particles completely protecting the tissues from any toxic effect.

In experiments by Belt and King (1943) and King, Rogers and Gilchrist (1945) powdered metallic aluminium mixed with quartz failed to check the development of silicosis. This failure to substantiate the claim of Denny, Robson and Irwin that silicosis can be prevented with aluminium may have been due to a loss of aluminium from solution in the lung, the quartz remaining in the tissues to exert its fibrogenic effect. In the experiments of Denny *et al.*, the animals breathed aluminium dust daily, and there was therefore a constant renewal of the aluminium in the lungs. In King's experiments the aluminium was given mixed with the quartz in a single dose by the intratracheal route (Kettle and Hilton, 1932).

The following results are from experiments which are a repetition of former work, but using aluminium hydroxide (Gardner, Dworski and Delahant, 1944) instead of metallic aluminium, and, in one experiment, a higher proportion of it, relative to the quartz, in the hope that there would be sufficient antidotal dust in the lung to remain during the life of the animal By this means the development of silicosis in rats was greatly retarded. A mixture of 90 per cent powdered quartz with 10 per cent powdered alumina produced only a slight reticulinosis, while quartz alone produced massive silicosis with numerous well-developed silicotic nodules. A parallel experiment in which quartz containing 1 per cent of admixed alumina was administered likewise produced only slight reticulinosis in ten months (Plate XI, Fig. 16 and Plate XI, Fig. 17).

It has been pointed out that aluminium can probably only exert its preventive or inhibitive action on silicosis production if it is constantly renewed in small amounts in the lungs. The only type of experiment where this can be easily accomplished is one where both quartz and aluminium are administered by daily inhalation. The experiments of King, Wright, Ray and Harrison, with a new kind of cabinet and a new method of producing

northern part of New York State. His colleague, Wright, devised and applied very thorough tests of respiratory function on the workers. Most of the men thought that their condition had improved, but it could not be shown for certain that this was the case. It was not possible to show any improvement on their X-ray films, nor to show that they had any statistically valid improvement of their respiratory function.

On the other hand, in many of the large plants where aluminium is being used, the men all think they are much better. They state that they can get their breath more easily, and whereas, before treatment, they found it difficult to walk upstairs, they now find that they can run up. This widespread conviction of subjective improvement was in contrast to the lack of any scientific demonstration that in fact much improvement, if any, had been effected.

In Denver, Colorado, a well-controlled experiment was conducted with men retired from the gold and copper mines. Aluminium hydroxide (instead of the metal) was used. Control subjects breathed fresh air through the same apparatus as was used to administer the aluminium hydroxide. There was the same subjective improvement here, unsupported by X-ray or physiological evidence, as seen in other centres. The controls were followed long enough to demonstrate that they also had subjectively improved, and Berry (1948) has published results which demonstrate that there was no valid evidence that alumina had improved the condition of the men taking the treatment over that of the controls.

There seems to be no doubt that silicosis can be prevented in animals by rendering the causative agent insoluble in the lung fluids, by coating the particles with aluminium. Simultaneous inhalation of aluminium with quartz prevents, or greatly retards, the development of silicosis; and the administration of aluminium subsequent to quartz has been claimed to arrest the progress of silicosis. Aluminium hydroxide, mixed with quartz, prevents the quartz from producing fibrosis. When administered to an animal with developed silicotic lesions, aluminium will retard their further development, and even cause some regression (Denny, Robson and Irwin, 1939). On the other

These consisted of collagen in the form of dense, rounded nodules with a few fibroblastic nuclei but no dust cells. They became larger and more numerous in the lungs of animals surviving to 400 days (Plate XII, Fig. 18).

Rats dusted with quartz and aluminium showed no lesions attributable to dust during the first 50 days. At about 200 days there were a few foci of dust cell reaction; but most of them were free from reticulin increase. Only in a few was there enough reticulin to warrant calling them 'nodules'; and these were much less dense and only about half the size of the nodules in the corresponding group of quartz-dusted rats. In rats living 300–400 days the lesions were less numerous than in the quartz-dusted rats, and they differed markedly both in size and density (Plate XII, Fig. 19). All of them were less than 500μ. in diameter, whilst the nodules due to quartz alone were almost double this size. Moreover, none of the nodules produced by quartz plus aluminium became more than mildly collagenous. The vast majority consisted only of a loosely woven tangle of fine reticulin fibrils. The difference in the size of the nodules was even more striking than the difference in the quality of fibrosis. Microincinerated sections showed that all the dust visible in ordinary sections (haematoxylin and eosin and silver) was siliceous; and that in individual nodules the concentration of dust was as high in the quartz plus aluminium lungs as in the pure quartz lungs.

Here, then, was a preventive, if not a cure, for silicosis. In Canada first and then in the U.S.A. aluminium has been tried as a prophylactic for silicosis. These experiments were first started at the McIntyre Mine, Ontario. Treatment of developed silicotics was attempted first, and careful medical, X-ray and physiological examinations were made. Unfortunately, the tests on a group of about 40 treated silicotics, begun early in the War, have not been repeated, as some of those responsible for the tests went into the Army. Consequently, no convincing answer has come from Canada as to whether the aluminium did these men any good.

Gardner, at Saranac Lake, conducted well-controlled experiments on zinc-miners, lead-miners and iron-workers in the

The four pure silica modifications all had very similar silica solubilities (about 14 mg./100 ml.) and size distributions, and were injected in equal amounts of 50 mg. corresponding to about 1,000 cm.^{2} of silica surface per rat. The resultant fibrosis varied considerably.

TABLE 4. Times for development of fibrosis with different forms of free silica

Form of silica	Minimum time (in months) required to produce fibrosis of grades 2–5			
	2	3	4	5
Fused silica	2	7	No further change	
Quartz	2	3	8	9
Cristobalite	1	2	7½	8
Tridymite	1	1½	–	2

With fused silica there was no further progress beyond grade 3 within seven to fifteen months. Quartz and cristobalite acted similarly. Grade 3 lasted from two or three months up to about seven or eight months and there was regular progression to grade 5. The cristobalite acted perhaps slightly faster than the quartz, as all grades were reached about a month earlier. The tridymite effect was spectacular. Grade 2 was reached after only one month, and grade 5 after two months. The fibrotic masses continued to grow, until in rats killed after twelve months there was hardly any normal lung tissue left (Fig. 21).

These results suggest that the crystal structure of pure silica *does influence the tissue reaction. Quartz, tridymite, cristobalite*, and fused silica are built up from silicon-oxygen tetrahedra of equal size and shape. The tetrahedra share oxygen corners only. Each silicon is surrounded by four oxygens and each oxygen is bound to two silicons. The feature that varies is the arrangement of the tetrahedra in space, and the bond angles which two silicons make with the oxygen that is bound to both. The tetrahedra can be packed tightly, as in quartz which has a density of 2·65, or more loosely as in the other modifications which vary in density from 2·2 for fused silica to 2·3 for cristobalite. The arrangement is regular with a three-dimensional

hand, no convincing evidence has been seen that aluminium will do the same in human silicosis. The most recent experiment, conducted by the Medical Research Council in this country, likewise appears to throw doubt on the usefulness of aluminium therapy or prophylaxis in the silicosis of pottery-workers (Kennedy, 1954). The solution of the silicosis problem probably lies in the suppression of dangerous dust, not in the introduction of other dust, even if that dust be antidotal.

Crystal Structure of Mineral Dusts in Relation to Silicosis

Pure silica (SiO_2) exists at room temperature in the four different modifications: quartz, tridymite, cristobalite, and fused silica. The first three are crystalline, but fused silica is a non-crystalline solid, a glass. Tridymite, cristobalite, and fused silica are metastable, but at room temperature they do not invert within measurable time to the only thermodynamically stable modification, quartz.

It can be regarded as proven that quartz can produce tissue changes in the lungs of animals similar to those found in human silicosis. It has also been thought that silica was unique in producing such changes. Experiments with *r*- and *l*-quartz (King, Rogers, Gilchrist, and Nagelschmidt, 1945) had shown that there was no difference in the action of either of the optically active forms from the effect of the racemic type.

Gardner (1938) injected rabbits by the ear vein with 1 g. lots of the four different modifications of silica and found progressive nodular fibrosis of the liver with quartz and fused silica, but a more severe diffuse fibrosis with cristobalite and a still more rapid reaction with tridymite which resulted in the death of all his animals within three months or less. King, Mohanty, Harrison and Nagelschmidt (1953a) injected fused silica, quartz, cristobalite and tridymite (Plate XIII, Fig. 20) of

($<3\mu$.) into the

e and severity of

. The results are

best summarized by listing the shortest time intervals required to produce the various grades of fibrosis.

If, therefore, any silicate can produce fibrotic lesions in the lungs of rats, felspar is the most likely to do it. Of the few experiments with felspar described in the literature the most recent one by Gardner (1938) showed no effect by intravenous injection into rabbits or intra-abdominal injection into guinea-pigs; and experiments by Mohanty *et al.* (1953) have shown that the introduction of potash felspar causes only the minimum of stage 1 fibrosis, even after seventeen months, although there was little, if any, evidence of dust removal. Thus, felspar is not even remotely as fibrogenic as quartz in this type of experiment.

Size of Dust in Relation to Pathogenicity

The importance of the size of the dust particles concerned in the production of silicosis has been widely appreciated, but the relationship between particle size and intensity of tissue reaction in man is complex. The size of the dust exerts a direct influence upon the number of particles suspended in the air to be inhaled, and on the penetrating power of the particles into the lung (cf. Davies, 1952; Watson, 1953). Other factors, such as distribution, weight, flocculation, and aggregation, and variation in the composition of dust in air, depend much on the size of the particles In the last few years the attention of many workers in this field has been drawn to investigating the range of sizes capable of entering the lung, and the relation of particle size of a particular dust to maximum pathogenicity In man it appears that it is the dust below 5μ. which is important in the production of lung disease. This field has been ably discussed by Hatch and Hemeon (1948) who point out the need for further research to determine the relation between toxicity of quartz and particle size.

As it has been possible to produce definite silicosis in experimental animals within a reasonably short period with free silica

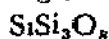
been carried out by Gardner and Cummins (1933) and Tebbens, Schultz, and Drinker (1945) Gardner and Cummins used intravenous injection into rabbits of quartz of diameter 1–3μ. and 6–12μ., and also of aluminium oxide 1–3μ. They found that

periodicity in the three crystalline modifications, and it is a random arrangement in fused silica. The crystal structures of cristobalite and tridymite are known in their main features but not in precise detail (Fig. 20).

It is not possible to say why the different modifications act so differently, and in particular what causes the very strong reaction of the tridymite. A closer examination of a number of

Fig. 21. Rates of development of fibrosis with different forms of free silica.

tridymite samples may give further information. But crystal structure alone cannot be the explanation of why some minerals produce severe silicosis and others do not, because the non-silicosis-producing felspars are very similar to quartz in chemical composition and crystal structure. Both are built up from three-dimensional networks of linked silicon-oxygen tetrahedra, and replacement of one out of four silicons of quartz by aluminium, accompanied by the introduction of one potassium or sodium to balance the charge, leads from the quartz to the felspar structure:


$SiSi_3O_8$
Quartz


$K,AlSi_3O_8$
Orthoclase felspar

actually being used up in the process, then surface also is the deciding factor. Therefore, the graded flint samples between <0·5 and 8μ. were given to rats by intratracheal injection both on the basis of equal weights (50 mg. per rat), and also on the basis of equal surfaces (700 cm.² per rat).

With equal weights the fibrogenic activity increased considerably with decreasing size. Whereas the largest fraction only gave stage 2 fibrosis in nine months, the smallest one gave stage 5 in five months. *With equal surfaces the largest fraction went only to stage* 3 and the smallest fraction to stage 4.

The main results can be summarized as follows:

TABLE 5. Constant-amount series (50 mg. flint per rat)

Flint size (μ)	Silica solubility (mg./100 ml.)	Surface per rat (cm.²)	Maximum stage of fibrosis	Reached after months	Minimum time (months) to give stage			
					2	3	4	5
4–8	5 1	170	2	9	9	–	–	–
2–4	8 9	350	3	10	5	10	–	–
1–2	13 1	750	5	9	1	5	7	9
0 5–1	13 4	2,000	5	7	1	2	3	7
<0 5	13 6	3,000	5	5	<1	<1	1	5

There was a regular increase in activity as the particle size decreased and the differences were large. The smaller the mean size of flint the faster or more advanced was the tissue reaction, when equal weights were injected.

The same samples given on an equal surface basis gave more homogeneous results when 700 cm.² of flint per rat gave different fibrosis stages after the following number of months.

TABLE 6. Constant-surface series

Flint size (μ)	Weight of dust per rat (mg.)	Silica solubility at constant area (mg /100 ml)	Minimum time (months) to give fibrosis stage		
			3	4	5
4–8	199	4 1	8	–	–
2–4	100	4 2	1	4	6
1–2	42	4 6	1	4	5
0 5–1	17 5	2 1	1	7	10
<0 5	8	4 1	1	10	–

the fine quartz fraction produced hyaline fibrosis, mainly in the liver, whereas the coarse quartz only gave a foreign body type reaction. The aluminium oxide did not produce any reaction.

Tebbens and his colleagues also used intravenous injection into the ear veins of rabbits of suspensions of 400 mg. quartz of mean sizes 3 3, 1·7, 1·0, and 0 6μ. diameter. These produced fibrosis in various organs, particularly the liver and spleen, and the smallest particles were much the most active. It was concluded that

> Particles of *1* micron and under are the most active and dangerous, not only in their ability to remain suspended in the air, but even of greater importance is their activity once they gain entrance to the deeper recesses and to the body and tissue.

Hatch and Kindsvatter (1947) made dust inhalation experiments on guinea-pigs with quartz below 0·5μ. diameter. The dusting period lasted from 4 to 20 weeks, and the authors calculated that up to 35 mg. per animal could have been retained in the alveolar tissue. Lung sections were illustrated showing fibrosis after 20 weeks, but no detailed pathological description was given. Attention was drawn to the rapidity of the lung changes, and it was claimed that 35 mg. per animal was

Fr

men·

dusts of a wide range of particle size, the upper limit of which should be less than 10μ. About the lower limit little is known; it is probably above 0·002μ. (cf King, 1947; Dale and King, 1953) and may be above 0·2μ To carry out further investigations on the relation of particle size to pathogenicity, and with dusts more accurately sized than those previously employed, King *et al.* (1953b) employed flint in several fractions covering the range from below 0·5μ to 8μ equivalent diameter

The pathogenicity of silica dust is probably a function of the total surface it exposes to the tissues. Two different mechanisms may be involved Solubility is certainly proportional to the surface and if the fibrogenesis is related to 'solubility' then a given dust will act in proportion to its surface. But, if it is not the dissolved silica but the surface itself that acts, perhaps as a catalyst promoting protein degeneration and fibrosis without

dusts, King (1945, ... this hypothesis ... found to be ... could be offered for this. ... of workers in pneumoconiosis research ... accepted the solubility theory of silicosis, some authors (Heffernan, 1935, 1946; Policard, 1947; Briscoe, 1949; Weyl, 1950) attribute pathogenicity to the freshly fractured surfaces of the silica particles. They suggest that 'the pathological process basically consists in hydration of the silica particle at the expense of the cell protoplasm'. They are also of opinion that freshly fractured particles have greater biological activity than old ones, and should be considered more pathogenic. Gardner (1938) and King (1945) could not find any difference in pathogenicity between freshly fractured quartz and quartz dust kept in the laboratory for many years in a dry state. With regard to the hydration of particles, Gardner (1938) conducted an experiment using quartz dust, kept in sterile physiological saline for several months, and found that it was as pathogenic as the dry dust. King (1947) suggested that the hydration of the respirable dust particles occurs immediately they are produced, or, at any rate, as soon as they enter the lung. Commenting on the 'Freshly Fractured Surface Theory' of silicosis, Wright (1950) has pointed out the need for further biological research to elucidate the pathogenic properties of the quartz dusts; and in recent German publications the contrast between the extremely toxic action of dissolved or colloidal silica as against the slow fibrogenic action of solid quartz has been stressed (Jotten, 1951; Holzapfel, 1952).

King (1947) found the apparent solubility of finely powdered quartz to decrease progressively from about 15 mg. dissolved SiO_2 per 100 ml. to less than 1 mg. when the quartz was repeatedly extracted with weekly changes of Ringer's solution of pH 7·4 (cf. Table 1); and in recent physiochemical studies Ritchie and Dempster (1952) have shown that quartz particles obtained by grinding contain a relatively highly soluble outer layer which can be removed by repeated buffer extraction and can be regenerated by grinding and in other ways. Under this layer is a relatively insoluble quartz core. This high-solubility

Since the largest fraction only just reached stage 3 and the smallest fraction stage 4, this series seemed to demonstrate that there was an optimum fibrogenic size in the region of 1 to 2μ. under the experimental conditions. If the failure of the fraction below 0·5μ. to produce stage 5 fibrosis can perhaps be explained on the ground that the absolute amount was insufficient, there is no simple explanation for the failure of the largest fraction to go beyond stage 3 (Plate XIV, Fig. 22).

The numbers of particles received by the animals in the different groups were grossly different. In the constant-surface experiment, despite the decreasing weight of dust used with decreasing size, the numbers of particles greatly increased. The rats receiving 199 mg. of 4–8μ. flint got about 500 million particles, those receiving 42 mg. of 1–2μ. got 10,000 million, and those receiving the 8 mg. of the smallest size (<0·5μ.) got 135,000 million. It might be contended that the largest-size dust (4–8μ.) failed to produce severe silicosis because of the small number of particles (if 500 million can be considered small), in spite of the large mass of dust, equal surface, and equal 'solubility'. But this can certainly not be claimed in respect of the smallest size (<0·5μ.) with which less-severe fibrosis, and less of it, resulted than with the 1–2μ. dust, despite there being enormously more particles, though less of weight but the same surface area and the same 'solubility'. As far as the constant-surface experiments are concerned, therefore, it seems justified to conclude that the finest particles of flint dust were less pathogenic than those in the 1–2μ. range which appears to point to there being a range of maximally fibrogenic sizes.

'Weathered' Dust

The importance of solubility of silica dust particles in causing silicotic lesions and tissue fibrosis has long been recognized and it has often been assumed that the pathogenicity of a silica dust would be directly proportional to its solubility. On the basis of the findings from an extensive study on solubility of different

hypotheses of fluoride adsorption or defect lattice on the basis of further experiments, and it is not impossible that our knowledge of the silicotic process may in this way be advanced.

Our conceptions about silica solubility and its relation to fibrogenesis may require modification. The previously found good agreement between solubility in the test tube and fibrotic effect in animals of pure quartz samples or rocks high in quartz, was due to the fact that solubility is related to the surface area of quartz per given weight. In mixture with aluminium or aluminosilicates, also, reduced solubility and delayed fibrogenesis go hand in hand. But it is now found that low solubility due to removal of the surface layer of quartz does not indicate less fibrosis. The resultant fibrosis is equal or it may even be higher. Despite the difficulties introduced by these new findings, however, a chemical theory of silicosis still appears to fit more of the known facts than does any other theory so far propounded. The conditions chosen for studying the 'solubilities' of siliceous dusts are arbitrary, and cannot adequately reflect the slow chronic leaching of mineral particles which goes on in the lungs. The high-solubility layer of the 'weathered' quartz may be restored in the lung, as it can be by grinding, or in the test tube by chemical treatment. Differences in crystal structure may imply differences in chemical as well as surface behaviour. Some recent experiments indicate that the initial rate of dissolution of tridymite is faster than that of cristobalite or quartz, which are slower than fused amorphous silica; and their surfaces may have different adsorption properties for cell constituents. A working hypothesis is useful, even though there may be difficulties in interpreting all the facts that come to light.

Acute Toxicity of Silica

Dale and King (1953) found that intravenous injections of 1–2 mg. colloidal silica caused rapid death in mice. Particulate silica of respirable dimensions (0·1–5μ.), and other dusts of similar size, were tolerated in similar amount, although large doses (30 mg.) caused death. The acutely toxic effect of colloidal silica was not abolished by the addition of aluminium hydroxide,

layer accounts for most of the silica 'solubility'. It is believed to be similar to the classical Beilby layer. Independently, Nagelschmidt *et al.* (1952) found evidence for an amorphous surface layer on fine quartz powders by quantitative determination of quartz by X-ray diffraction methods, before and after etching the powders with hydrofluoric acid.

Animal experiments were conducted by King *et al.* (1953c) to test the pathological significance of these observations. Two sets of quartz samples were prepared, (*a*) by repeated leaching with Ringer's solution, and (*b*) by etching with hydrofluoric acid. The untreated quartz samples served as controls.

Stripping with Ringer's solution reduced the silica solubility to 10 per cent of its original value, but had no effect on the fibrosis, which reached stage 5 with sample and controls equally in 5–6 months. Etching with hydrofluoric acid, however, greatly enhanced the fibrogenic reaction.

These experiments were of a preliminary character and no

results with HF-extracted quartz where a sample with silica solubility of 1·2 mg. SiO_2/100 ml. acted faster than its control with a solubility of 9·1 mg. In addition there was a specific effect in the HF-extracted series which causes extremely rapid fibrogenesis. It is possible that this specific effect was due to fluoride or HF adsorbed on the surface of the quartz. The fact that fluoride could not be demonstrated chemically may indicate an exchange of F' for OH' on the quartz surface, but this requires further study. An alternative explanation of the additional effect would be the following. as a result of the rapid etching with HF, there was a defect lattice with a high lattice energy and highly fibrogenic properties. Washing with saline or electrodialysis caused a rearrangement of the surface of the crystals to a more stable structure with less fibrogenic properties. It may be possible to decide between the alternative

area than particulate silica of 'respirable' dimensions. But in the absence of any knowledge of what 'surface' of silica particles implies in the way of special properties, this seems to beg the question. It may be that particles of silica, particularly those of the crystalline varieties, have special surface properties which may cause biological effects independent of any silicic acid that may dissolve away from them. It could, for instance, be argued that the prevention of silicosis by coating quartz particles with aluminium can be explained in this way, although the prevention of any silica going into solution, because of the coating with aluminium, seems a more probable explanation.

It is interesting that substances which absorb strongly on quartz dust, e.g. aluminium, dyes and proteins, have no effect on the acute toxic properties of colloidal or 20 Ångstrom silica, as well as having none on quartz itself, although one of them (aluminium) so markedly lessens the fibrogenic activity of quartz.

Possible Mechanism of Silicosis

Although these results have no direct bearing on silicosis they do suggest a possible mechanism of silicosis production. Fine quartz particles release silicic acid into solution in a liquid or in the liquid contents of a cell. Some of the silicic acid escapes into the blood, and is carried away to be excreted; other silicic acid molecules polymerize to colloidal dimensions, and become fixed on the protoplasm of the cell. Colloidal silicic acid is a protein precipitant (Rappaport and Reifer, 1937), and its action in precipitating some of the protein of the engulfing cell may be the cause of the typical fibrotic response which is always caused by the presence of free silica particles in tissues. The quartz particles, however, must be of fairly definite size range. They must not be too small, like 20 Ångstrom silica, or they will produce an acute toxic effect. If they are too big, no appreciable dissolution will take place. The formation of silicotic nodules is a slow chronic phenomenon. It appears that to produce them silica particles must be small enough to liberate slowly, and over a long time, soluble silica into solution from their surfaces in sufficient amount to produce a chronic local irritative effect; and must not be so small that they will quickly

dyes, or protein; substances which are strongly adsorbed on silica of larger particle size.

The colloidal silica was more than 10 times as toxic as quartz of larger dimensions. Quartz varying in size from 0·2 to 8μ. showed no marked difference in toxicity, nor did any other of the siliceous materials tried. Various quartzes, flint, kaolin, aluminiums and coals, all of particle sizes from about 0·1 to 10μ., appeared to be rather similar in the amounts of them required to kill. There is, however, an apparent break between their required amounts and those of 20 Ångstrom (0·002μ.) silica and colloidal silicic acid. With the big drop in size of particles, the silica became very much more lethal; and it seemed probable that the acutely toxic, 'pharmacological' effect of silica was a matter of particle size and given only by suspended silica of colloidal dimensions. The much larger amounts of larger particulate silica required to kill may well have contained some colloidal matter in their suspensions (cf. Clelland, Cumming and Ritchie, 1952).

It is apparent that not all foreign colloidal matter is toxic in this way; colloidal carbon (Indian ink), various colloidal dyes and colloidally dispersed metals and oxides of metals can, as is well known, be tolerated in large amounts. Thus, King and Dolan (1933) injected 3 ml. per kilogram of Thorotrast (609 mg. thorium dioxide) intravenously into rabbits without any apparent effect, either acute or chronic; and Indian ink has been injected into mice in amounts of 0·2 ml. (equivalent to 22 mg. carbon) before signs of weakness appeared. Silica may not be unique, but we do not know of any other substance which is highly toxic only in the colloidal state, while remaining relatively very much less toxic at a larger size, and apparently non-toxic when adsorbed in a molecularly disperse state, as is done constantly in large amounts by all herbivorous animals which ingest silica in their food, absorb a lot of it, and promptly excrete it in the urine (King and Stantial, 1933). The reverse is, of course, quite common, e.g. dissolved arsenic is lethal, while colloidal arsenic can be tolerated in considerable amount.

It could, of course, be argued that the phenomenon is entirely due to surface area, i e. that colloidal silica has a much greater

CLELLAND, D. W., CUMMING, W. M. and RITCHIE, P. D. (1952). *J. app. Chem.* 2, 31.
DALE, J. C. and KING, E. J. (1953). *Arch. indust. Hyg. Occ. Med.* 7, 478.
DAVIES, C. N. (1952). *Brit. J. indust. Med.* 9, 120.
DENNY, J. J., ROBSON, W. D. and IRWIN, D. A. (1939). *Canad. Med. Ass. J.* 40, 213.
DIMERBROCK (1649). Cited by Schumann, S. *Industrial Dust. Its control with special reference to legal measures.* Quoted by Thackrah, 1832. p. 85.
GARDNER, L. U. (1923). *Amer. Rev. Tuberc* 7, 344.
GARDNER, L. U. (1938). *Amer. Inst. Min. & Met. Eng. Tech. Pub.* 929.
GARDNER, L. U. and CUMMINS, D. E. (1933). *Amer. J. Path* 9, 751.
GARDNER, L. U., DWORSKI, M. and DELAHANT, A. B. (1944). *J indust. Hyg.* 26, 211.
GREENHOW, E. H. (1865). *Trans. path. Soc. Lond.* 16, 59.
GREGORY, J. C. (1831) *Edinb. med. J.* 36, 389.
GYE, W. E. and KETTLE, E. H (1922). *Brit. J. exper. Path.* 3, 241.
GYE, W. E. and PURDY, W. J. (1922). *Brit. J. exper. Path.* 3, 76.
HALDANE, J. S. (1917). *Trans. Inst. Eng.* 55, 264.
HALDANE, J. S. (1931). *Trans. Instn. Min. Eng. Rep.* Annual Gen. Meeting.
HATCH, T. and HEMEON, W. C. L. (1948). *J. indust. Hyg.* 30, 172
HATCH, T and KINDSVATTER, V. H (1947). *J. indust. Hyg.* 29, 342.
HEFFERNAN, P. (1935). *Tubercle*, 16, 397.
HEFFERNAN, P. (1946). *Brit med. J.* 1, 928.
HOLT, P. F and OSBORNE, S. G. (1953) *Nature*, 171, 892
HOLZAPFEL, L. (1952) *Beitrage zur Silikose-Forschung*, Heft. 15, p. 1.
HIPPOCRATES (460–370 B C.). *Medical works of Hippocrates* translated Chadwick, J. and Mann, W. A. (1950)
JOTTEN, K W. (1946) *Schr. a. d Hyg Inst. d. Univ Münster 1*
JOTTEN, K. W. (1951). *Beilage zum Bundesarbeitsblatt*, No. 10, Oct. 1951. (Saf Mines Res Est Rep Translation No. 3490.)
JOTTEN, K. W and GARTNER, H. (1951). *Med. Welt*, 44, 1.
KENNEDY, M. (1954). To be published.
KETTLE, E H (1932). *J Path. Bact.* 35, 395
KETTLE, E H. and HILTON, R. (1932) *Lancet*, 1, 1190.
KING, E J. (1928). *J. biol. Chem* 80, 25
KING, E. J. (1939). *Biochem J.* 33, 944
KING, E. J. (1945). *Med Res. Coun Spec Rep* Lond No. 250.
KING, E J. (1947) *Occ Med.* 4, 26.
KING, E J and DOLAN, M (1933). Unpublished
KING, E J, MOHANTY, G P., HARRISON, C V. and NAGELSCHMIDT, G. (1953a) *Brit J indust Med.* 10, 9
KING, E. J MOHANTY, G P, HARRISON, C. V. and NAGELSCHMIDT, G. (1953b) *Brit J indust Med* 10, 76
KING, E. J MOHANTY, G. P, HARRISON, C. V. and NAGELSCHMIDT, G (1953c) *Arch indust Hyg Occ Med* 7, no. 6, p. 455

and completely dissolve and be either carried away or produce only an acute local effect, which will lack the continuous stimulus of constantly released colloidal silicic acid. It may be this prolonged irritative, or stimulating, effect of constantly released colloidal silica which is responsible for the formation of the fibrous scar tissue of a silicotic nodule (cf. Jotten and Gartner, 1951 and Holt and Osborne, 1953, for recent similar ideas).

Belt (1939) described the structure of the silicotic nodule as a core of densely packed fibres containing numerous doubly refractive particles of silica, surrounded by a sheath of fibres (ball-of-twine structure). Micro-incineration revealed a dense halo of fine silica at the periphery, but a space almost vacant of silica between the periphery of the fibrous sheath and the core (cf. Fig. 5). This structure strongly suggests a diffusion, outward from the central deposit of particles, of silicic acid, much of which becomes fixed on protoplasmic constituents in a zone slightly outside and surrounding the core, where micro-incineration reveals it as a multitude of very fine amorphous particles, easily seen in dark-field illumination, but not visible [illegible] or colloidal silica.

ACKNOWLEDGEMENT

I AM INDEBTED TO THE EDITOR OF THE *British Journal of Industrial Medicine* FOR PERMISSION TO REPRODUCE SEVERAL OF THE PHOTOMICROGRAPHS.

REFERENCES

AGRICOLA, Georgius (1494–1555). *De Re Metallica*, Book I Translated from first Latin edition of 1556 by Herbert C Hooer and L. H Hoover. Mining & Scientific Press, San Francisco, 1912.

BELT, T. H (1939). *J Path. Bact* 44, 39

BELT, T. H and KING, E J. (1943) *J Path. Bact* 55, 69

BELT, T. H. and KING, E J (1945) *Med. Res. Counc Spec Rep* Lond No 250

BELT, T. H, FERRIS, A A and KING, E J. (1940) *J Path. Bact* 51, 263

BERRY, J. W. (1948). *Amer. Rev Tuberc.* 57, 557

BREMNER, D. (1939). *Canad. Min. J.* Oct

BRISCOE, H V. A (1949). *Chem. Trade J.* 124, 478

VIII

The Principles of Ganglionic Block

W. D. M. PATON

In this lecture I would like to reverse the usual logical order of such a talk and to start by describing briefly some of the practical applications of ganglion blocking agents before discussing the pharmacological principles involved in their use. For the position is now somewhat as follows: it was first shown in the laboratory that such drugs were suitable for clinical trial; then these clinical trials proved sufficiently promising to make it worth carrying on with therapeutic investigations, and at the same time raised a number of new problems. As a result, whereas two or three years ago the laboratory worker was asking the clinician to try this or that substance, now it is the clinician who turns to the laboratory worker for interpretation of his results.

The main uses of hexamethonium, which I shall take as the prototype of ganglion blocking agents for the moment, have been in the treatment of *hypertension*, in *gastroenterology*, and in the *reduction of bleeding* at operation. Thus, in a patient suffering from hypertension, his blood pressure may be greatly reduced, his enlarged heart get smaller, his pulmonary oedema regress, and his vision, impaired by retinal damage, improve. Similarly in peptic ulcer hexamethonium produces a reduction of acid secretion and a depression of the exaggerated motility of the gut. The third use mentioned is in some ways the most interesting. If the blood pressure is reduced during a surgical operation to levels of about 60–70 millimetres systolic, bleeding can be enormously lessened. This was done in the past either by giving the patient a deliberate haemorrhage, or by inducing spinal anaesthesia.

King, E. J. and McGeorge, M. (1938). *Biochem J.* **32**, 417.
King, E. J., Rogers, N. and Gilchrist, M. (1945). *J. Path. Bact.* **57**, 491.
King, E. J. and Roman, W. (1938). *Biochem. J.* **32**, 2283.
King, E. J. and Stantial, H. (1933). *Biochem. J.* **27**, 990.
King, E. J., Wright, M., Ray, S. C. and Harrison, C. V. (1950) *Brit. J. indust. Med.* **7**, 27.
Knight, A. J. (1830). *North of England Med. & Surg. J.* **1**, 174
Kussmaul, A. (1930). *Silicosis Records of Internat. Conf.* Johannesburg, p 339.
Marshall, W. (1834). *Lancet*, **2**, 271.
Mavrogordato, A. E. (1926). *Contributions to the study of miners' phthisis. Publ. S. African Inst. for Med. Res.* No. 19.
Miller, A. and H. (1869). *My Schools and Schoolmasters.* Edinburgh.
Mohanty, G. P , Roberts, D. C., King, E. J., Harrison, C. V. and Nagelschmidt, G (1953). *J. Path. Bact.* **65**, 501.
Nagelschmidt, G., Gordon, R. L. and Griffin, O. G. (1952). *Nature*, **169**, 539.
Paracelsus, K. F. (1493–1541). *Ges Gewerbehyg. N.F.* **12**, 1925.
Peacock, T. B (1861). *Trans. path Soc. Lond.* **12**, 36.
Policard, A. (1933). *Presse Med.* **41**, 89; *La Med du Trav.* **5**, 171.
Policard, A. (1947). *Inst. Min Eng. & Inst. Min. Metal* Lond. p. 15
Ramazzini, B (1700). Cited by Collins
Rappaport, F. and Reifer, I (1937). *Mikrochim. Acta.* **1**, 20.
Ritchie, P. D. and Dempster, P B. (1952). *Nature*, **169**, 538
Sosman, R. B (1927). *J. Amer. Chem. Soc* (monograph).
Tebbens, B. D., Schulz, R. Z and Drinker, P. (1945). *J. indust. Hyg.* **27**, 199.
Thackrah, C T. (1832). *The Effects of Arts, Trades and Professions.* Longmans, 2nd edit
Thomson, W. (1837). *Med Chir. Trans* Lond. **20**, 230
Titus, A. C. (1937). *J. indust. Hyg. Tox* **19**, 138.
Visconte, -- (1870). Reported by Rovinda, C. L , 'Un Case di Silicosi del Pulmone con Analist Chimica'. *Analli di Chimica*, 1871.
185
Wright, B. M. (1950) *Nature, Lond.* **166**, 538
Zenker, F A. (1865). *Silicosis Records of Internat. Conf.* Johannesburg (1930). p 339.

The electrical activity accompanying a nerve impulse in the preganglionic trunk causes, when it reaches the terminations of the preganglionic nerve, a discharge of acetylcholine (Feldberg and Gaddum, 1934). This impinges on the ganglion cell body, which appears to be specifically differentiated so that it responds to this acetylcholine by a vigorous electrical discharge. This is sufficient, in turn, to excite a further impulse in the post-ganglionic nerve trunk of the same type as that which, in the preganglionic trunk, initiated the whole chain of events. In normal transmission the cycle of events is complete in a fraction of a second, and the electrical activity of the ganglion has completely subsided within 100–400 milliseconds (Eccles, 1935). This is because acetylcholine, as soon as it has achieved its excitation of the ganglion cell body, is then either destroyed by the cholinesterase located at the preganglionic nerve terminals, or diffuses away from the site of action, to be rapidly diluted in the surrounding tissue space.

Because the chain of transmission involves a chemical link at this point, it is vulnerable to a chemical attack. First we may have substances bearing a chemical relationship to acetylcholine sufficient to give them some affinity for the ganglion cell membrane, but lacking the additional ability of actually initiating excitation. Such compounds, it is believed, then simply compete with acetylcholine for the ganglion cell receptors, thus rendering acetylcholine less effective than usual. If these compounds are present in sufficient concentration they may reduce the effect of acetylcholine below that required to initiate impulses in the postganglionic trunk, so that ganglionic transmission fails. A block of this kind, which is exerted by drugs such as d-tubocurarine, T.E.A. and hexamethonium, is termed a *competitive* block, and these drugs are *competitive* blocking agents.

But another type of block is also possible. A drug may actually be able to mimic acetylcholine completely but to be unsusceptible to the action of cholinesterases. In that case it excites the ganglion for much longer than the fraction of a second involved in normal transmission and this is accompanied by a prolonged *depolarization* of the ganglion cell membrane (Paton and Perry, 1953). With this prolonged excitation of the postganglionic

But both of these methods have considerable disadvantages, and it has proved possible to achieve the same end by giving a dose of hexamethonium and then tilting the patient so the blood may accumulate in the lower part of his body. When this technique is successful, very dry fields of operation can be obtained, which make plastic surgery much easier and quicker, and neurosurgery a great deal simpler; sometimes, in fact, the use of controlled hypotension may make possible operations otherwise almost too hazardous to undertake.

But each of these uses has raised problems. In *hypertension* we have the difficulties that oral dosage is relatively inefficient, that patients vary enormously in their response, and that patients on hexamethonium for a relatively prolonged period develop an increasing resistance to its hypotensive action. Similarly in the use of hexamethonium in *gastro-intestinal disease*, where it has been used to depress the motility of the small intestine, it has

to diminish *bleeding* at operation raises the question of why it is that a patient may have his pressure lowered so far without suffering any ill-effects, and the additional practical question of how to make it possible to achieve these reductions of blood pressure more uniformly.

These are the sort of problems which arise, and others will no doubt appear in the future; and it is with such questions in mind that we must turn to consider the sort of knowledge needed to be able to handle ganglion blocking agents with some hope of success. This may be summarized briefly as knowledge about (1) the properties of the transmission process in autonomic ganglia; (2) the properties inherent in quaternary salts; (3) normal autonomic background tone and reflex activity; (4) the reactions of the body to autonomic paralysis.

1. The Mechanisms of Ganglionic Transmission and Block

The autonomic ganglia receive from the central nervous system preganglionic trunks which enter the ganglia, and there synapse with the cells from which the postganglionic trunks take origin.

shown in animals and in man to synergize with hexamethonium in producing ganglionic block (Mason and Pelmore, 1953). It indicates a potentially valuable way of supplementing the action of a competitive ganglion blocking drug.

TABLE 1. Types of Ganglionic Block

Characteristics	Competition with acetylcholine	Depolarization	Prevention of acetylcholine release
Release of acetylcholine	Normal	Normal	Depressed
Response to injected acetylcholine	Depressed	Depressed	Present
Initial action	Block	Stimulation	Block
Examples	d-tubocurarine T.E.A. (tetra-ethyl-*ammonium*) Hexamethonium (C 6) Nicotine (2nd phase)	Acetylcholine T.M.A. (tetra-methyl-*ammonium*) Nicotine (1st phase)	Local anaesthetics Calcium deficiency Botulinum toxin

Ganglion blocking agents of the competitive type are at present the most widely used. Those acting by depolarization are not very serviceable clinically because of the initial stimulation which they produce; this means, for instance, that before being able to lower the blood pressure with a drug like nicotine, a phase of strong autonomic excitation, with a rise of blood pressure and release of adrenaline, would have to be undergone. Among the competitive blocking agents hexamethonium has so far proved most generally suitable and most of this review will refer to information gained with this substance in the clinical studies which it has made possible. But many other substances have been prepared, progeny still very like their parent in appearance, but sometimes more active and longer lasting. (No doubt hexamethonium will in due course be completely superseded; but the lessons learnt with it are likely to be of general application.) Another competitive ganglion blocking agent is tetraethylammonium, with which interest in ganglion blocking agents in recent years was first aroused (Acheson and Moe,

nerve trunk, there results finally a condition where acetylcholine still impinges on the ganglion cell membrane normally, but this is already maximally excited, cannot be further excited, and transmission fails. A block of this sort by nicotine or acetylcholine itself, which is always associated with an initial stimulant effect, may be termed depolarization block.

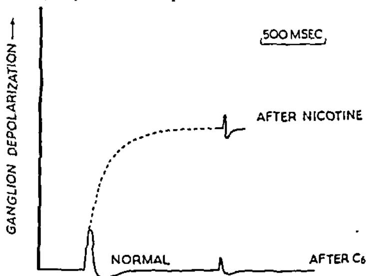
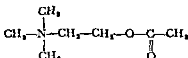


FIG. 1. Diagram of normal ganglion action potential and action potential (a) after *nicotine*, (b) after *hexamethonium*. With *nicotine* the ganglion is depolarized, as though a normal excitation had persisted, with hexamethonium the ganglion is not depolarized.

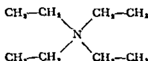
A third type of intervention with ganglionic transmission may arise in a different way. It is known that some substances, including the local anaesthetic procaine and its stable analogue procaine amide, can interfere with the release of acetylcholine from the preganglionic nerve terminals (Harvey, 1939; Paton and Thompson, 1953). This is not associated with any chemical resemblance of procaine to acetylcholine, and the 'pharmacological lesion' is of a different kind to the others already mentioned; it may be attributed to a particular susceptibility to local anaesthetics on the part of the delicate nerve terminals. A development of particular interest is the clinical combination of procaine amide with hexamethonium. The former has been

TABLE 2. The Structural Formulae of some Ganglion Blocking Agents

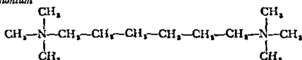
Acetylcholine



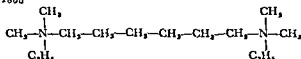
Tetraethylammonium



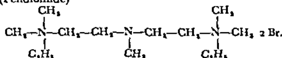
Hexamethonium



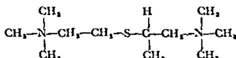
M & B 1863



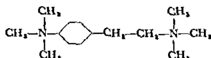
Ciba 9295 (Pendiomide)



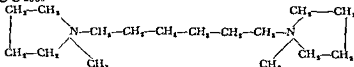
Roche: 3–0484



M & B 1950



M & B 2050



1945–6); it still has a place in investigation for its transience of action. But, as we shall see, its action is complicated in some degree, which detracts distinctly from its value.

The use of hexamethonium is not in practice, however, completely straightforward. This is not because the drug itself produces irregular effects when given under standard conditions to a particular autonomic ganglion; on the contrary, perfectly consistent results can be obtained in animals with almost any of these drugs. It arises because autonomic activity in man is a rather more subtle affair than was suspected. The first two complications are that ganglia differ in their sensitivity to competitive block; and that the intensity of the block is related to the frequency of excitation of the ganglia.

(i) *Selectivity and Maximality of Block.* It has in fact been known for many years that ganglia differ in their susceptibility to paralysis, whether by nicotine, d-tubocurarine or (more recently) hexamethonium. Indeed difference in sensitivity may occur between the different ganglion cells within one particular ganglion, e.g. the superior cervical ganglion. Here cells supplying the nictitating membrane seem to be resistant compared to

sensitivity of a ganglion directly, and it cannot be assumed that a given dose adequate to block one particular system would also prove sufficient to block another. Some capriciousness in the incidence of block is already clear in man. For instance, paralysis of accommodation only occurs in a proportion of subjects. Some subjects develop a paralysis of the light reflex without effect on the reaction to accommodation—a pharmacological Argyll-Robertson pupil. Similarly paralysis of the bladder, or constipation, may be absent or trivial, despite the appearance of considerable reduction in blood pressure or postural hypotension.

Related to this is the observation that it is rarely possible to produce complete ganglionic block throughout the body. In particular some of the vasomotor ganglia in animals appear to

so effectively protects the central nervous system from quaternary salts. If this were the case, the surgical and pharmacological inaccessibility would spring from the same cause—that the ganglion cells concerned still bore too close a relationship (anatomically or structurally) to the central nervous mass from which they derived.

(ii) *Fatigability*. One of the notable features of paralysis by ganglion blocking agents is the increased ease of fatigue of transmission which ensues. This may be shown in various ways.

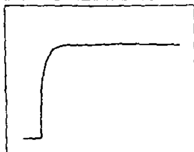




FIG. 3 Record of contraction of the cat's nictitating membrane, excited by stimulation of the preganglionic cervical sympathetic trunk at 10 shocks per second for one minute Before hexamethonium and after hexamethonium, showing onset of 'fatigue'.

First, the block may be greater as the rate of excitation of the ganglion increases Second the effect of a given dose of hexamethonium becomes greater, the longer excitation has been established. One may see the onset of fatigue from the earliest moment of a period of excitation (Fig. 3). The response of the effector organ (in this case the nictitating membrane) starts quite vigorously but then wanes within a few moments until quite a considerable degree of block has occurred. A similar phenomenon has been observed in man After a moderate dose of hexamethonium, and in an individual in which this produces some interference with accommodation, the subject frequently notices that during recovery from the effects of the drug he can read successfully for half a minute or a minute, and then finds his vision becoming blurred again. A brief rest from the effort of

be very resistant to complete paralysis. This probably means that in man complete blockade will be difficult to achieve, although no doubt in particular individuals certain autonomic functions can be completely suppressed. In clinical experience it has been found that, at any rate in the lower limb, hexamethonium will produce as great an increase in blood flow as will a block of the autonomic pathways with local anesthetics. On the other hand it has proved very difficult to suppress entirely such vascular reflexes as the 'cold pressor response'.

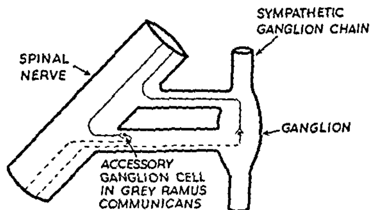


FIG 2. Diagram and accessory ganglionic synapse in grey ramus communicans Continuous lines show preganglionic trunks, dotted lines the postganglionic trunks.

There is an interesting analogy here with the failure of surgical sympathectomy to be complete. This has been shown to be due to the existence of a considerable number of sympathetic ganglion cells outside the main paravertebral trunks (Boyd and Monro, 1949); they are found, for instance, in the rami, and in the spinal nerves, positions often inaccessible to the surgeon. Fig. 2 illustrates a typical situation, where the preganglionic trunk synapses in a grey ramus, and the postganglionic trunk dodges back (as if anticipating surgical assault) into the spinal nerve. There may be more than a superficial analogy here. These accessory ganglion cells might be regarded as being still incompletely extruded from the central nervous system, and could still be partly enclosed by the blood-brain barrier which

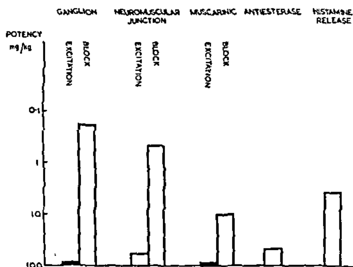


Fig 4 (*a*). Spectrum of pharmacological actions possessed by *d-tubocurarine*. The ordinate gives the approximate potency of the drug (in mg /kg.) for each action. Only the more prominent actions are represented, and each compound (*see also b, c, d*) has a number of other properties in addition.

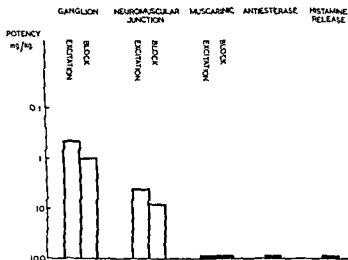


Fig 4 (*b*). Spectrum of pharmacological actions possessed by *nicotine*.

accommodation enables him again to see, for a short time, with clarity.

This enhanced fatigability of the partially blocked ganglion may be important clinically. For it follows that the ganglion subjected to the most intense autonomic drive may be also the ganglion most readily blocked by the drug. Thus in neurogenic hypertension or in peptic ulcer, the abnormal drive on the ganglia concerned may be preferentially weakened before other, normal, functions are interfered with. This would explain how, occasionally, very small doses (1–5 mgms.) of hexamethonium may, when first administered to a hypertensive patient, produce a very considerable fall in blood pressure.

2. The Properties of Quaternary Salts

(i) *Other Pharmacological Actions.* Besides being transmitter at the ganglionic synapse acetylcholine mediates two other transmissions at least: the peripheral parasympathetic transmission and transmission across the neuromuscular junction. Further, enzymes capable of destroying it at a very high rate are widely distributed at important points throughout the body. Most substances bearing a pharmacological relationship to acetylcholine at any one of these various points also have traces of a relationship at the other sites. One may thus construct a sort of spectrum of the activity of the drug. At each site it may imitate acetylcholine or antagonize it, or antagonize the enzyme which destroys acetylcholine. Thus quite a diverse range of actions in mimicking, blocking, or preserving the natural transmitter at the various synapses have to be considered. In addition, since these compounds are nitrogenous bases they are liable to possess other actions, such as that of histamine release. A spectrum of this kind has been constructed for *d-tubocurarine* (Fig. 4*a*) to illustrate the point. It will be seen that it possesses detectable activity at all the synapses mentioned, as well as some anticholinesterase action and the power of histamine release. *Nicotine* (Fig. 4*b*) is likewise fairly complex. With *tetraethylammonium* (Fig. 4*c*) also, although it is considerably better than these in the selectivity of its action, there are several additional effects at other sites. The fourth spectrum represented is that of

ACTIONS RELATED TO ACETYLCHOLINE

GANGLION | NEUROMUSCULAR JUNCTION | MUSCARINIC | ANTIESTERASE | HISTAMINE RELEASE | PARAESTHESIAE etc.

POTENCY mg/kg

EXCITATION BLOCK EXCITATION BLOCK EXCITATION BLOCK

0·1

1

10

100

Fig. 4(*c*). Spectrum of pharmacological actions possessed by *tetraethyl-ammonium*.

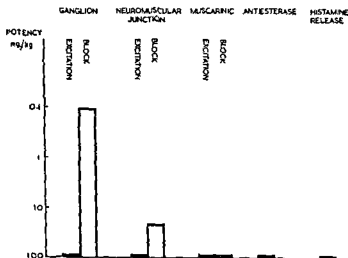


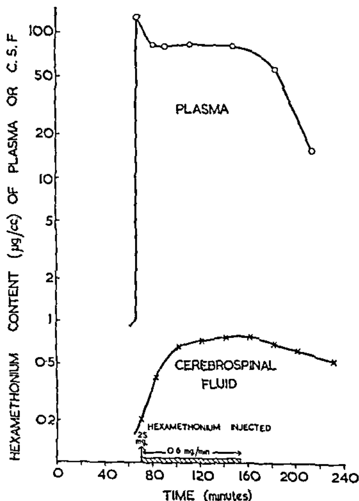
Fig 4 (*d*). Spectrum of pharmacological actions possessed by *hexamethonium*.

hexamethonium (Fig. 4*d*) and illustrates its relative pharmacological cleanness (Paton and Zaimis, 1952).

The significance of a spectrum such as this lies in the use made of such drugs. It does not matter if a drug has, for instance, some atropine-like action as well as a ganglion paralysing action, if its use is to be mainly in the treatment of peptic ulcer; but investigations with such a drug will always be difficult to interpret and will throw little more light on what is happening, than can investigations using a mixture of atropine and hexamethonium. In a similar way, the ability of tetraethylammonium sometimes to elicit a secretion of adrenaline or noradrenaline, has complicated the interpretation of the results obtained with it and led possibly to some mistaken conclusions. Again, an awareness that all these actions are liable to occur in agents able to block ganglia, allows more careful scrutiny of the pharmacological background to such drugs before they are applied to man. Finally—a scientific point—the properties of a drug give information about the receptor structure at which it acts: each can be regarded as a sort of garment, whose design gives one a clue to the structure it clothes. Nicotine could be compared to one of those shapeless objects that people knit in railway trains —one has a grave doubt whether it is a baby's vest, or a seaman's stocking. But hexamethonium is more like a glove, enabling one to say that it goes on something with five fingers (or rather that it fits on to something about 7 Å wide). In our poverty-stricken state, where everyone talks about receptor structures but no one has seen one, we must be grateful even for these small clues about their nature.

(ii) *The Pharmacodynamics of Quaternary Salts.* The relationship of ganglion blocking agents to acetylcholine shows itself in another way. In order for a drug to be able to antagonize acetylcholine at the ganglionic synapse, it is necessary for it to be chemically related to acetylcholine. This implies that it must be a very strong base. This in turn entails that it shall penetrate cell membranes very slightly. This means that a drug such as hexamethonium has to lead a predominantly extra-cellular existence. It follows that it is only slowly absorbed from the intestine since there is here a complete cellular barrier between

individuals, and having eliminated all the variables of distribution and excretion, the response of the organism to a ganglion blocking drug is widely different from one individual to another. What other factors are at work? Two at least may be



the lumen of the intestine and the blood stream. When absorbed it is distributed roughly in the extra-cellular space (Morrison and Paton, 1953). While present in the blood stream it penetrates only very slightly across the cell barrier of the choroid plexus and appears only in traces in the cerebral spinal fluid (Fig. 5) (Paton, unpublished). It penetrates only slightly into the aqueous humor and into the red cells. Finally it is excreted only by glomerular filtration, and like inulin can neither traverse the renal tubules outwards by being secreted nor yet inwards by being re-absorbed.

The effects of a dose of hexamethonium therefore are determined by, amongst other factors, its absorption from the intestine, by the volume of the extra-cellular fluid and by the glomerular filtration rate of the subject concerned. Because intestinal absorption is poor and (even worse) erratic, parenteral administration is now the rule. As the drug is eliminated almost exclusively by the kidney, prolonged actions are obtained in patients with seriously impaired kidney function.

An Analysis of Variations in Sensitivity. From what has been said, it might be expected that if one took a single test (say the effect of hexamethonium on the standing systolic blood pressure), in a normal subject, then the results might vary according to the distribution of the drug in the body, but that they would relate pretty well with the plasma concentrations. A recent investigation was made of this point (Morrison and Paton, 1953), in which effects of subcutaneous injections of hexamethonium on the standing systolic blood pressure were compared with the concentration in the plasma (assayed by biological means).

It was found that subjects did indeed vary considerably both in the rate of absorption from the subcutaneous site, in the volume of distribution of the drug, and in its excretion. But these factors did not account for all the differences by any means, and considerable differences between individual responses were obtained even when the effects were plotted against the plasma concentrations. It turned out, in fact, that they differed as much as eightfold in the size of the cardiovascular response to a given plasma hexamethonium level.

Thus, even when using a single autonomic function, in normal

under which there is undoubtedly a vigorous activity in these pathways. Among these conditions will be many reflexes, of which a number of vascular reflexes are known to be considerably affected by hexamethonium. For instance, if one makes a violent expiration against a resistance—the so-called Valsalva manœuvre—when it is over there is quite a considerable rise in blood pressure, often called an 'over-shoot'; hexamethonium greatly reduces it. Another vascular reaction is that to haemorrhage, which is normally followed by a compensatory peripheral vasoconstriction such that a pint of blood can be withdrawn without producing more than a trivial fall of blood pressure. In the presence of hexamethonium, which prevents much of this response, the effects of haemorrhage on the blood pressure become much greater (Fig. 6) (Freis *et al.*, 1951); and the circulatory system almost behaves like a simple elastic tube in which the pressure falls as fluid is withdrawn from it and rises again when it is restored. This is an important observation because bleeding may be more dangerous in a patient treated with hexamethonium than in the normal.

A third reflex of great practical importance is the postural reflex (Fig. 7). Normally when a subject stands, the blood pressure does not fall very much and sometimes even rises slightly. To bring this about, considerable vascular re-adjustments are necessary; the blood flow diminishes to almost all organs except to the head; the cardiac output diminishes and the venous filling pressure falls. After hexamethonium this fairly wide vasoconstriction no longer takes place, or to a much reduced degree. The cardiac output is diminished on standing about as much as in the normal. Hence the arterial pressure must go down. This postural reaction to hexamethonium, due to interference by the drug with normal postural tone, is of primary importance. Its mechanism is not completely understood, and we do not know how far it arises simply because peripheral arteriolar constriction can no longer occur, or how far it is due to the accumulation of blood in the legs or abdomen, or elsewhere. We know that the blood pressure can be raised again by exercise, by standing-running, or by abdominal compression. The reaction is exploited deliberately in the treatment of

envisaged; variations in background autonomic tone, and variations in the response to partial autonomic paralysis.

3. Autonomic Tone

The effect of a ganglion blocking agent, if this is its only significant action, is entirely indirect. When there is no natural activity traversing the autonomic pathways, then block of the

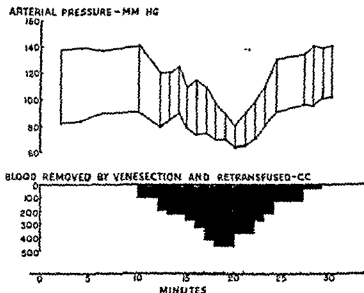


Fig. 6. Effect of withdrawing 500 cc. blood in successive steps, and then replacing it, on the blood pressure of a supine human subject after injection of 50 mgm. hexamethonium intravenously. The fall of blood pressure is roughly proportional to the amount of blood removed, and the loss of only 100 cc. is sufficient to lower the pressure significantly. (From Freis *et al.* (1951))

ganglionic transmission in that pathway will not produce any physiological effect. For instance, in an animal whose nervous system has been completely destroyed, so that there is no longer any autonomic outflow, the administration of hexamethonium has no action at all. Our knowledge of normal tonic activity therefore becomes very important. Dr. Feldberg[1] describes in his paper the sort of pathways that exist and some of the conditions

[1] *See* 'The Physiology of the Autonomic Nervous System', pages 343–47 below.

under which there is undoubtedly a vigorous activity in these pathways. Among these conditions will be many reflexes, of which a number of vascular reflexes are known to be considerably affected by hexamethonium. For instance, if one makes a violent expiration against a resistance—the so-called Valsalva manœuvre—when it is over there is quite a considerable rise in blood pressure, often called an 'over-shoot'; hexamethonium greatly reduces it. Another vascular reaction is that to haemorrhage, which is normally followed by a compensatory peripheral vasoconstriction such that a pint of blood can be withdrawn without producing more than a trivial fall of blood pressure. In the presence of hexamethonium, which prevents much of this response, the effects of haemorrhage on the blood pressure become much greater (Fig. 6) (Freis *et al.*, 1951); and the circulatory system almost behaves like a simple elastic tube in which the pressure falls as fluid is withdrawn from it and rises again when it is restored. This is an important observation because bleeding may be more dangerous in a patient treated with hexamethonium than in the normal.

A third reflex of great practical importance is the postural reflex (Fig. 7). Normally when a subject stands, the blood pressure does not fall very much and sometimes even rises slightly. To bring this about, considerable vascular re-adjustments are necessary; the blood flow diminishes to almost all organs except to the head; the cardiac output diminishes and the venous filling pressure falls. After hexamethonium this fairly wide vasoconstriction no longer takes place, or to a much reduced degree. The cardiac output is diminished on standing about as much as in the normal. Hence the arterial pressure must go down. This postural reaction to hexamethonium, due to interference by the drug with normal postural tone, is of primary importance. Its mechanism is not completely understood, and we do not know how far it arises simply because peripheral arteriolar constriction can no longer occur, or how far it is due to the accumulation of blood in the legs or abdomen, or elsewhere. We know that the blood pressure can be raised again by exercise, by standing-running, or by abdominal compression. The reaction is exploited deliberately in the treatment of

hypertension and in controlled hypotension in anaesthesia; but it represents a complication in other therapeutic uses such as the treatment of peptic ulcer.

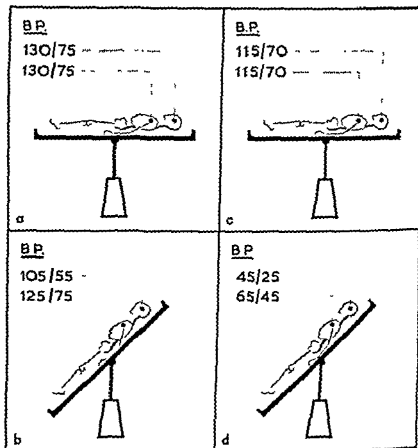


FIG. 7. Diagram of the effect of change of posture on the blood pressure before (a and c) and after (b and d) hexamethonium.

But in all these cases there are obvious stimuli at work (*raised intrapulmonary pressure, haemorrhage, posture*), and it is relatively easy to relate the action of the drug to the reflexes involved. It is a rather different matter to assess how far under normal undisturbed conditions there is a continuous autonomic tone. For instance, in an audience sitting listening to a speaker,

how far are the viscera or the skin or the heart-rate being influenced by autonomic impulses of various kinds? We know the paths are there, but are they being used? One of the uses of hexamethonium may be, perhaps, that it will tell us a good deal about these paths. To illustrate the sort of information that it should be possible to obtain, I have compiled a list of actions which have already been described in patients receiving hexamethonium (Table 3). One can infer from these observations that there is for instance a steady neurogenic lubrication of the mucous surfaces (i.e. the conjunctiva, the nose and the mouth, and throat), and of the skin and scalp. They illustrate, too, how the normal vasoconstrictor tone of the skin must be an important sluice damming back heat within the body; and they bring to light a tone in the conjunctival vessels not commonly recognized.

The Hexamethonium Man. To summarize what I have said about the background autonomic tone, I would like to be frivolous for a moment and to bring together all the actions described, supposing them all exerted in one individual, an entirely hypothetical individual, whom we could call the 'hexamethonium man'. He is a pink-complexioned person, except when he has stood in a queue for some time, when he may get pale and faint. His handshake is warm and dry. He is a placid and relaxed companion; he may laugh, but he can't cry because the tears cannot come. Your rudest story will not make him blush and the most unpleasant circumstances will fail to make him turn pale. His wife will testify that his collars and socks stay very clean and sweet. He wears corsets and may, if you meet him out, [illegible]
vascul[illegible]
legs). [illegible]
moisten his dry mouth and throat. He is rather long-sighted and easily blinded by bright light. The redness of his eye-balls may suggest irregular habits and in fact his head is rather weak. But he always behaves like a gentleman and never belches nor hiccups He tends to get cold and keeps well wrapped up. But his health is good; he does not have chilblains and those diseases of modern civilization, hypertension and peptic ulcer pass him by. He is thin because his appetite is modest; he never

feels hunger pains and his stomach never rumbles. He gets rather constipated so that his intake of liquid paraffin is high. As old age comes on frequency, precipitancy and strangury will not worry him, but he will suffer from retention of urine and impotence. One is uncertain how he will end, but perhaps, if he is not careful, by eating less and less and getting colder and colder, he will sink into a symptomless, hypoglycaemic coma and (like the universe) die a sort of entropy death.

TABLE 3. Effects of Hexamethonium in Man: I

SIGN OR SYMPTOM	AUTONOMIC PATHWAY BLOCKED
	Sympathetic supply (adrenergic)
Injection of conjunctivae	to conjunctival vessels
Flushing of skin and capillary pulsation	to arterioles of skin
Rise in skin temperature	,, ,,
Rise in skin blood flow	,, ,,
Rise in limb blood flow (leg arm)	to skin and muscle vessels and veins (?)
Reduction of cold pressor response	to arterioles and veins (?) of skin, muscle and viscera
[illegible]	,, ,,
[illegible]	,, ,,
[illegible]	to retinal vessels
[illegible]	to arterioles of skin
Relief of causalgia	to arterioles of affected region
Sensitization to insulin	to adrenal medulla
	Sympathetic supply (cholinergic)
Dry scalp and dry skin	to sweat glands
Reduction of reflex sweating	,, ,,

This is a caricature, but it was worth perpetrating to emphasize the point that it is crucial to know what our background autonomic tone is; and that it is already possible to glean a good bit of information about it from clinical reports already available.

4. REACTIONS TO GANGLIONIC BLOCK

A final element in the complexity of the response to ganglionic block must be considered. If one functionally removes a large part of the autonomic nervous system, it is only to be expected that some sort of reaction, in an attempt to compensate for it, will take place. It was mentioned earlier that in patients treated

for hypertension, a tolerance to hexamethonium steadily develops. It never seems to become absolute, and usually falls in blood pressure can still be obtained if a big enough dose is given. The tolerance passes away again if the treatment is stopped and can be re-established if treatment is resumed. The question arises whether this is due to a generalized ganglionic tolerance or whether it is something specific to the blood pressure. There is definite evidence that so far as certain para-

TABLE 3. Effects of Hexamethonium in Man. II

SIGN OR SYMPTOM	AUTONOMIC PATHWAY BLOCKED
	Parasympathetic supply (cholinergic)
Paralysis of ocular accommodation and reaction to light	Ciliary ganglion
Dryness of eyes	Sphenopalatine ganglion
Dryness of mouth and nose	Otic ganglion and chorda tympani
Dryness of larynx	Vagal ganglia—in larynx
Tachycardia	Vagal ganglia—in heart
Diminished motility of stomach and small intestine	Enteric ganglia
Constipation	,, ,,
Ileus	,, ,,
Diminished secretion of gastric juice:	
spontaneous	,, ,,
nocturnal	,, ,,
in response to insulin	,, ,,
Depression of bladder reflexes	Pelvic nerve ganglia
Impotence	,, ,, ,,

sympathetic ganglia are concerned, this tolerance either does not develop or develops much more slowly (Morrison, 1953). In a group of patients in whom the dose of hexamethonium has had to be steadily increased in order to produce a roughly constant effect on the blood pressure, the incidence of paralysis of accommodation, of the bladder, and of the intestine steadily rises. It is possible that there is a slight accommodation of this sort affecting all ganglia, but it is certain that such accommodation is not the whole story. It is necessary to look elsewhere for an explanation, and so far none is yet available.

A more usual reaction to a fall in blood pressure is, of course, an increased activity of certain autonomic pathways. It is worth

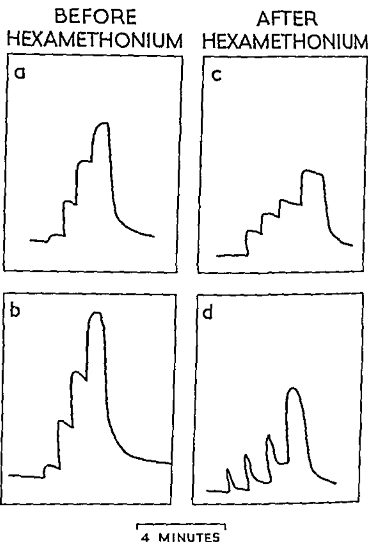


Fig. 8. Records of contraction of the cat's nictitating membrane excited by stimulation of the preganglionic cervical sympathetic trunk at 10 (a and c) and at 20 (b and d) shocks per second before and after hexa-

[illegible]

total postganglionic discharge to be increased.

making a distinction here between what can be referred to briefly but loosely as 'recruitment' and 'acceleration'. One could envisage an intensification of autonomic activity in two ways, eit [illegible] inactive, [illegible] already e[illegible] the cat's nictitating membrane excited through the cervical sympathetic with the superior cervical ganglion as a relay station between the two. In Fig. 8 is reproduced a rather crude recruitment obtained by increasing the stimulus strength to the nerve fibre in four steps. In the normal preparation this brings successive increases in activity of the nictitating membrane. On the other side of the figure the same thing has been done after hexamethonium. As one would expect, although the ganglia rapidly fatigues, nevertheless the recruitment of successive fibres does actually confer some benefit. So that it may be said that in the presence of ganglionic block by hexamethonium a successful reaction could take place by recruitment, provided that the block was not so intense as to paralyse completely these newly activated synapses as well, or so intense that they fatigued rapidly to insignificant levels, when the reaction would only be a *transient one.*

But the second kind of response, by *acceleration*, is rather a different story. In Fig 9 are the results of two experiments on a ganglion that is partially paralysed. At the top are the contractions which the ganglia will normally produce to excitations of one, ten and twenty shocks a second. When the ganglion is partially paralysed, an acceleration of rate of discharge from one a second to ten a second does in fact produce some increase in the contraction height. But the acceleration from ten a second to twenty a second actually weakens the final contraction We have therefore a situation in which the benefit obtained by increasing the rate of autonomic discharge is rather severely limited. Of course there is a similar 'ceiling' in normal ganglia where you can produce fatigue by stimulating at very high frequencies The characteristic effect of hexamethonium is not the establishing of a ceiling but of bringing it down to comparatively low levels of stimulation rate.

M

One important application of this may be when the reaction involves the release of adrenaline. Under normal conditions the autonomic activation of the adrenals appears to be negligible. It follows that, when activity is initiated, the preganglionic discharge rate is likely to be quite low, still well below its 'ceiling', and thus resistant to hexamethonium. Thus the apparently paradoxical situation could readily arise in which a vigorous adrenal discharge could take place in the presence of ganglionic block by hexamethonium sufficiently intense to lower the blood pressure enough to excite the adrenal response. Dr. Morrison and Dr. Enderby tell me that they have quite often observed responses to hexamethonium which they were strongly tempted to believe were of this kind. One may compare this with what might be expected from the general vasoconstrictor fibres in a case of hypertension. One must assume that in this patient, supposing him to be of the neurogenic type, there is a considerable activity in the vasoconstrictor fibres already. If in such a patient the blood pressure is lowered by hexamethonium, reaction to this by increasing the rate of autonomic discharge will have little effect, or may even lead to a more intense paralysis (as in Fig. 9).

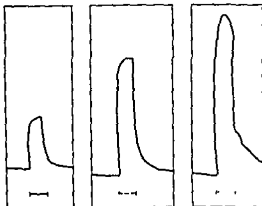
These are rather theoretical considerations. But with care it might be possible to use them to illuminate clinical practice. In general, one might expect that those reactions to the effects of ganglionic block which are fresh and are not normally active will be more successful than reactions involving intensification of activities habitually in play.

Conclusion

The general trend of the argument has been to multiply reasons for the variation of the actions of hexamethonium in man; reasons such as differences in sensitivity of various ganglia, in

[illegible]

list and lends weight to the occasional clinical complaint that such drugs are unreliable in action. Yet, under standard condition in anaesthetized animals, they are extremely consistent in





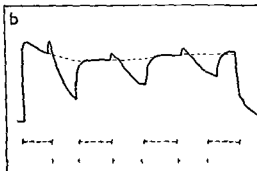


FIG 9. Records of contraction of the cat's nictitating membrane excited by stimulation of the preganglionic cervical sympathetic trunk at 1, 10 or 20 shocks per second (denoted by continuous, broken and dotted lines respectively). Above are the responses to these rates of stimulation, for one minute, in a normal cat. Below are the responses obtained after hexamethonium.

Increasing the rate of stimulation in the presence of hexamethonium does not necessarily lead to a bigger postganglionic response, although the change from 1/second to 10/second in (a) does so, the further acceleration from 10/second to 20/second in (b) actually produces (after an initial augmentation) a considerable increase in block.

One important application of this may be when the reaction involves the release of adrenaline. Under normal conditions the autonomic activation of the adrenals appears to be negligible. It follows that, when activity is initiated, the preganglionic discharge rate is likely to be quite low, still well below its 'ceiling', and thus resistant to hexamethonium. Thus the apparently paradoxical situation could readily arise in which a vigorous adrenal discharge could take place in the presence of ganglionic block by hexamethonium sufficiently intense to lower the blood pressure enough to excite the adrenal response. Dr. Morrison and Dr. Enderby tell me that they have quite often observed responses to hexamethonium which they were strongly tempted to believe were of this kind. One may compare this with what might be expected from the general vasoconstrictor fibres in a case of hypertension. One must assume that in this patient, supposing him to be of the neurogenic type, there is a considerable activity in the vasoconstrictor fibres already. If in such a patient the blood pressure is lowered by hexamethonium, reaction to this by increasing the rate of autonomic discharge will have little effect, or may even lead to a more intense paralysis (as in Fig. 9).

These are rather theoretical considerations. But with care it might be possible to use them to illuminate clinical practice. In general, one might expect that those reactions to the effects of ganglionic block which are fresh and are not normally active will be more successful than reactions involving intensification of activities habitually in play.

Conclusion

The general trend of the argument has been to multiply reasons for the variation of the actions of hexamethonium in man; reasons such as differences in sensitivity of various ganglia, in absorption, distribution and excretion of the drug, in the existence of autonomic tone and in its intensity, and in the reactions to ganglionic block which may take place. It makes a formidable list and lends weight to the occasional clinical complaint that such drugs are unreliable in action. Yet, under standard condition in anaesthetized animals, they are extremely consistent in

IX

Cholinesterases and Anti-Cholinesterases

R. H. S. THOMPSON

THE mechanism by which acetylcholine is enzymically destroyed inside the body is a process which in recent years has acquired special importance owing to advances made in a number of different branches of biology. In the field of toxicology, for example, the introduction of the anti-cholinesterase insecticides, which on account of their high toxicity to man are now presenting us with a problem in industrial medicine, and the related so-called 'nerve-gases' serve to stress the importance of an understanding of cholinesterase action and of the effects of its inhibition. In liver disease and in malnutrition it has been shown that the level of the serum cholinesterase is diminished, while in the field of anaesthetics certain of the new muscle relaxants present us with the special problems of the properties of acetylcholine and cholinesterase at the motor end-plates.

The problem of the kinetics of the cholinesterase group of enzymes, and of their inhibition by the various anti-cholinesterases, has recently been reviewed (Whittaker, 1951; Nachmansohn and Wilson, 1951). It is proposed in this article to deal

tion by anti-cholinesterase compounds, and lastly to summarize what evidence we have concerning the physiological function of this group of enzymes.

their behaviour. But if one considers the situation in man, it is perhaps not very surprising that such variation exists. The autonomic nervous system pervades the whole body; it carries all the play of emotions and nervous influences into their visceral expression; and everyone differs in these as much as in their physical structure. The issue might therefore be put in a different way; that there is, in our present wealth of somewhat uncontrollable responses and of unanswered questions, a really rich field for greatly increasing our knowledge of normal and pathological autonomic function.

REFERENCES

ACHESON, G. and MOE, G. K. (1945–6). *J. Pharmacol.* **84**, 189; **87**, 220.

BOYD, J. D and MONRO, P. A. G. (1949). *Lancet*, ii, 892.

ECCLES, J. C. (1935). *J. Physiol.* **85**, 464.

FELDBERG, W. and GADDUM, J. H. (1934). *J. Physiol.* **81**, 305

FREIS, E. D., STANTON, J. R , FINNERTY, F. A., SCHNAPER, H. W., JOHNSON, R. L., RATH, C. E. and WILKINS, R. W. (1951). *J. Clin. Invest.* **30**, 435.

HARVEY, A. M (1939). *Bull Johns Hopk. Hosp.* **65**, 223

MASON, A. A. and PELMORE, J. F. (1953). *Brit. Med. J.* i, 250.

[illegible]

PATON, W. D. M. and ZAIMIS, E J. (1952). *Pharmacol. Rev.* **4**, 219.

Rudney, 1943). At about the same time Nachmansohn and Rothenberg (1945) classified the cholinesterases into two corresponding types which they called 'specific' and 'non-specific' cholinesterase respectively.

Just as selective substrates were discovered for these types of esterase, so selective inhibitors soon began to be described. Although there is now a large series of such compounds the best known and most exhaustively studied inhibitor of pseudo-cholinesterase is di*iso*propyl fluorophosphonate (DFP), whose action was first described by Adrian, Feldberg and Kilby (1947); percaine (Zeller and Bissegger (1943)) and the synthetic prostigmine analogue Nu683 (the dimethylcarbamate of 2-hydroxy-5-phenylbenzyl-trimethylammonium bromide) studied by Hawkins and Gunter (1946) were also early described as selective inhibitors of this enzyme. Caffeine (Zeller and Bissegger, 1943), bis-β-chloroethyl-N-methylamine, or 'nitrogen mustard' (Thompson, 1947; Adams and Thompson, 1948), and another prostigmine analogue Nu1250, N-*p*-chlorophenyl-N-methylcarbamate of *m*-hydroxy-phenyl-trimethylammonium bromide (Hawkins and Mendel, 1949), were among the first selective inhibitors of true cholinesterase to be described.

The selectivity of most of these compounds for one enzyme or the other is only relative, and in most cases higher concentrations of the inhibitor produce a significant inhibition of the other type of cholinesterase also. But by the use of selective substrates and inhibitors it is possible to classify the cholinesterase activity of most mammalian tissues into one or other of Mendel's two main types.

Mendel's classification has not, however, escaped criticism, particularly by Nachmansohn. While much of this criticism is concerned largely with nomenclature, more recent work has shown that the criteria by which Mendel appears to have originally defined his two types of cholinesterase are not in every case fulfilled, and that, judged by these criteria, there would seem to be more than two types of cholinesterase. For example, in 1946 Mazur and Bodansky demonstrated that mouse brain can hydrolyse triacetin more rapidly than acetylcholine and that this hydrolysis of triacetin is highly sensitive to inhibition

The Classification of the Cholinesterase Group of Enzymes

The cholinesterases are distinguished from the other esterases present in animal tissues by their property of splitting choline esters more rapidly than non-choline esters. It was soon realized, however, that there is more than one type of cholinesterase. Alles and Hawes in 1940 showed that the cholinesterase in human erythrocytes differs from that in human serum with regard to the speed with which acetylcholine is split at varying substrate concentrations, the enzyme in red cells being much more active at low concentrations of acetylcholine (i.e. the Michaelis constant is small). They also showed that the cholinesterases in these two tissues differed in their substrate specificities, acetyl-β-methylcholine being hydrolysed only by the erythrocyte cholinesterase. Further evidence for the existence of at least two different types of mammalian cholinesterase came from the work of Glick (1941), Richter and Croft (1942), and Zeller and Bissegger (1943). Mendel and his colleagues in Toronto next proposed a general classification of the cholinesterases into two main types which they termed 'true' and 'pseudo' cholinesterase. The term 'true' cholinesterase was given to the enzyme present in the erythrocytes of many species and in the nervous system; this enzyme was at that time thought to be active only against choline esters and to be partially inhibited by high concentrations of acetylcholine. The term 'pseudo' cholinesterase, on the other hand, was applied to the enzymes present in serum and pancreas which hydrolyse both choline and non-choline esters, although splitting the former more rapidly than the latter, and which follow the usual Michaelis relationship in showing maximal activity at high acetylcholine concentrations (Mendel and Rudney, 1943a). These workers also employed 'selective' substrates for differentiating between the two types of cholinesterase by using acetyl-β-methylcholine, which, as indicated above, is readily hydrolysed by the true enzyme but not by the pseudo-cholinesterase, and benzoylcholine which is hydrolysed by the latter enzyme but is not attacked by the true cholinesterase (Mendel, Mundell and

of rat heart, which according to its sensitivity to selective inhibitors is a pseudo-cholinesterase, hydrolyses propionylcholine more rapidly than either acetyl or butyrylcholine.

It would seem in fact that the cholinesterases of mammalian tissues comprise a group of related enzymes in so far as their substrate specificity is concerned, and that on this criterion alone it is not possible to divide them into only two groups.

Allowing for these differences in substrate specificity, however, the cholinesterases of mammalian tissues would seem to fall into two main groups as regards their sensitivity to inhibitors and as regards the physiological effects of their inhibition *in vivo*. Mendel's classification into true and pseudo-cholinesterases is still, in the author's view, perhaps the most satisfactory, in that less is implied in this terminology than in the other proposed nomenclatures, and what is implied is largely physiological; thus, while there is considerable evidence that the true cholinesterase present in nervous tissue and in skeletal muscle is concerned with inactivating acetylcholine, since signs referable to acetylcholine accumulation develop when this enzyme is inhibited, the case for relating the pseudo-cholinesterase to acetylcholine destruction, particularly in connection with innervated structures, is much less secure. Although, as Nachmansohn has pointed out, there are obvious objections to the use of the prefix 'pseudo' in this connection, it is a valuable reminder of our present state of uncertainty as to the physiological role played by this enzyme, and is, by virtue of its drawbacks, clearly only a temporary designation.

THE DISTRIBUTION OF CHOLINESTERASES IN VERTEBRATE TISSUES

A knowledge of the distribution of true and pseudo-cholinesterases in the different tissues of the body is desirable before one can hope to arrive at any over-all conclusions as to their functions. Although there are still many gaps to be filled in, we now have, for certain animal species, fairly detailed information as to their occurrence and localization

As regards the cholinesterases in blood, the types present vary somewhat from species to species. In general the erythrocytes contain true cholinesterase, although Augustinsson (1948) has

by anti-cholinesterases, even though mouse brain had been regarded as containing only true cholinesterase which, according to the criteria then existing, should be active only against esters of choline. In the same year Bodansky (1946) also showed that the true cholinesterase in human erythrocytes, which had been freed from the non-specific esterase, could nevertheless still hydrolyse triacetin. It is now also clear that the idea of 'selective' substrates for the differentiation of the different types of enzymes is only relative, and that the enzymes present in the tissues of different animal species vary widely in their behaviour towards these substrates. Thus, Levine and Suran (1950) have reported that the cholinesterase in swine serum is inactive towards both acetyl-β-methylcholine and benzoylcholine, while Earl and Thompson (1952a) have shown that the cholinesterase of chicken plasma hydrolyses both of these so-called selective substrates.

In an attempt to classify the cholinesterases more exactly in terms of their substrate specificity, Adams and Whittaker (1949) studied in a systematic way the effect of variations in the structure and configuration of the substrate. They showed that both true and pseudo-cholinesterases are 'unspecific' in that they both hydrolyse non-choline esters, although, as Whittaker (1951) has pointed out, there is a marked preference in the case of both enzymes for these aliphatic substrates which approach most closely the configurations of choline esters. The idea that pseudo-cholinesterases are less specific than true cholinesterases receives therefore no support from these specificity studies. Their findings with esters containing different acyl groups did, however, suggest an underlying difference between these two enzyme types, in that the true cholinesterases studied by them showed optimal activity against acetates (e.g acetylcholine), whereas butyrates (e.g. butyrylcholine) were hydrolysed most rapidly by the pseudo-cholinesterases; they therefore proposed the names 'aceto-cholinesterase' and 'butyro-cholinesterase' respectively for the two types. Unfortunately, further work with cholinesterases from a wider range of animal sources shows that this classification also is not generally applicable; for example, Ord and Thompson (1951) have shown that the cholinesterase

(Thompson and Whittaker, 1944), but its activity in the presence of selective substrates had not been studied.

From these results it would appear that, by contrast with brain, skeletal muscle and the adrenals, those tissues in the rat which receive only post-ganglionic autonomic innervation contain either a very significant amount of, or a preponderance of,

TABLE 1. Relative activities of True and Pseudo-cholinesterases in Rat Tissues
(Acetyl-β-methylcholine and benzoylcholine as substrates)

Tissue	Activity in presence of selective substrate, expressed as % of rate of acetylcholine hydrolysis	
	True	*Pseudo*
Brain	75	6
Diaphragm	54	0
Quadriceps femoris	69	7
Adrenal gland	39	2
Stomach	20	24
Liver	20	23
Lung	21	25
Salivary gland	33	31
Heart ventricle	9	29
Heart auricle	9	32
Intestinal muscle	18	44
Intestinal mucosa	7	50
Harderian gland	6	52
Skin	9	25

the pseudo-cholinesterase. In this species therefore it is interesting to note that those tissues in which acetylcholine exerts a nicotine-like action (i.e. in which it is concerned with transmission to another neurone or to a striated muscle cell) chiefly contain the true cholinesterase, while those in which it exerts a muscarine-like action contain a relatively large amount of pseudo-cholinesterase. Whether this apparent correlation is of physiological significance, however, is not yet known, and more work is needed in other species before any final conclusion can be drawn

At about the same time Koelle (1950) studied the distribution of cholinesterases in the cat, using a histochemical approach; he

reported that fowl red corpuscles do not show any cholinesterase activity. Human and horse serum contain chiefly pseudo-cholinesterase, although the sera of certain other species (for example, the guinea-pig and the rat) contain both types. In the red cells the cholinesterase appears to be bound to the cell membrane, and can be precipitated with it.

Apart from work on these enzymes in blood, the first fact to emerge as regards the other tissues of the body was that the predominant cholinesterase of the brain resembles that in the erythrocytes (Mendel and Rudney, 1943a; Nachmansohn and Rothenberg, 1945), and may therefore be classified as a true cholinesterase. In 1944 Langemann showed that human skeletal muscle contains only this latter type. Gradually, however, evidence accumulated showing that pseudo-cholinesterases were also widespread in different tissues, for example, in dog pancreas (Mendel and Mundell, 1943), human ovary (Langemann, 1944), peripheral nerve fibres and ganglia of the cat (Sawyer [illegible]

[illegible]

survey of the distribution of cholinesterase types in rat tissues, using selective substrates, summation experiments and selective inhibitors to identify the enzymes. They found significant amounts of pseudo-cholinesterase in all the tissues studied except diaphragm. It appeared from their work (Table 1) that rat tissues could be classified into three groups depending on the relative amounts of true and pseudo-cholinesterase present:

Group A. Tissues containing largely or entirely true cholinesterase (brain, skeletal muscle and the adrenal glands).

Group B. Tissues containing both enzymes in amounts such that acetyl-β-methylcholine and benzoylcholine are hydrolysed at about the same rate (stomach, liver, lung and salivary gland).

Group C. Tissues containing chiefly pseudo-cholinesterase (heart, intestinal muscle and mucosa, Harderian gland and skin).

The pseudo-cholinesterase contents of heart (both auricle and ventricle) and intestine were strikingly high. The presence of an active cholinesterase in skin had been described earlier

after it was reported from a number of sources that whole brain preparations can in fact hydrolyse benzoylcholine, although at a

tain areas benzoylcholine is hydrolysed relatively rapidly. Ord

TABLE 2. Distribution of Cholinesterases in Grey and White Matter of Cerebrum

(Acetyl-β-methylcholine and butyrylcholine as substrates)

Species	Activity in presence of selective substrates, expressed as % of rate of acetylcholine hydrolysis	
Grey matter	*True*	*Pseudo*
Human	53	43
Rabbit	81	24
Rat	75	32
Guinea-pig	62	23
Cat	98	5
Dog	59	65
White matter		
Human	19	206
Rabbit	31	117
Rat	37	50
Guinea-pig	22	266
Cat	23	194
Dog	46	181

and Thompson (1952) have examined this benzoylcholine-hydrolysing ability of the brain in more detail, using both benzoyl and butyrylcholine (Nachmansohn and Rothenberg, 1945; Adams, 1949) as substrates, and comparing also the sensitivity of the hydrolysis of different choline esters to inhibition by a number of selective inhibitors. All the areas of the human brain which were studied were found to contain a pseudo-cholinesterase. Whereas the basal ganglia, cerebellum, thalamus and cerebral cortical grey matter contained chiefly the true cholinesterase, the subcortical white matter exhibited considerably more pseudo-cholinesterase activity than true. This was also found to be so in the cerebral white matter from a number of animal species (Table 2).

was mainly concerned with the localization of these enzymes in the nervous system, but he found that in the cat the liver contains almost exclusively the pseudo-cholinesterase, and in the ileum, although as in the rat both types are present, pseudo-cholinesterase predominates in the longitudinal muscle layer and in the muscularis mucosae.

Although the pharmacological actions of acetylcholine on blood vessels have been extensively studied, the cholinesterase activity of isolated arteries has only recently been examined. A study of a number of different rabbit and human arteries and of the aorta of the rabbit, rat and guinea-pig showed, however, that they are relatively rich in cholinesterase, which, in the species so far examined, has in each case been a pseudo-cholinesterase (Thompson and Tickner, 1953); the aorta in particular is rich in this enzyme; in the case of the rat, for example, butyrylcholine is hydrolysed at a rate of about 20,000 $\mu l. CO_2/g./hr.$, the highest rate which we have so far found in any tissue.

Distribution of Cholinesterases in the Nervous System

In view of the role of acetylcholine in transmission processes, and the part played by the different cholinesterases in acetylcholine metabolism, the distribution and localization of these enzymes in the central and peripheral nervous systems is of particular importance.

Nachmansohn, to whom we are perhaps mainly indebted for our present knowledge of the role of cholinesterases in nerve conduction and transmission, was the first to compare the ability of different areas of the brain to hydrolyse acetylcholine, and to demonstrate that the grey matter of the brain is more active in this respect than the white (Nachmansohn, 1939). Mendel and Rudney (1943b) examined the brains of a number of vertebrate species, and concluded that brain contains only the true cholinesterase and no pseudo-cholinesterase; Nachmansohn and Rothenberg (1944) also concluded that brain contains only the 'specific' cholinesterase. On the other hand, both types of enzyme were found to be present in peripheral nerve (Boell, 1945; Sawyer and Hollinshead, 1945; Sawyer, 1946), and soon

Acute and Chronic Effects of Anti-Cholinesterases

The potent and relatively specific inhibitors of cholinesterase that have been introduced since the last war have provided us with valuable tools for studying the properties of these enzymes both *in vitro* and *in vivo*. Amongst the best known of these new organo-phosphorus anti-cholinesterases are di*iso*propyl fluorophosphonate (DFP) and tetra-ethyl pyrophosphate (TEPP).

The acute effects produced by the injection of these compounds into animals are all compatible with the view that they are due to an accumulation of acetylcholine in the tissues resulting from inhibition of cholinesterase activity. These effects have been described in detail in a number of publications, and it is not proposed to list them here.

More recently, however, it has been found that certain of the anti-cholinesterases can cause chronic demyelinating lesions in the nervous system. The compound that has been most extensively studied from this point of view is tri-*ortho*-cresyl phosphate. Medical interest in this substance was first aroused when it was shown to be the causative agent of an outbreak of motor neuritis in the United States (Smith and Elvove, 1930; Smith, Elvove and Frazier, 1930). The histopathology of the paralysis, both in man and experimental animals, has been described by Smith and Lillie (1931), who found evidence of demyelination both in the peripheral nerves and in the white matter of the spinal cord, together with degenerative changes in the anterior horn cells. Bloch (1941), who first showed that tri-*ortho*-cresyl phosphate (TOCP) is an inhibitor of cholinesterase, suggested that the paralysis might be due to accumulation of excessive amounts of acetylcholine at the motor end-plates, resulting from inhibition of the true cholinesterase at that site.

However, in a brief report Mendel and Rudney (1944) stated that although the serum cholinesterase of the rat is lowered by this compound, the true cholinesterase is unaffected. As most workers have found that paralysis and demyelination do not occur in the rat following the administration of TOCP, Earl and Thompson (1952a) studied the *in vitro* action of TOCP on esterases from human and hen tissues (both sensitive species);

These findings suggest that pseudo-cholinesterase in the central nervous system is associated particularly with the myelinated fibre tracts. From their work on dog brain Burgen and Chipman (1951) have also concluded that in this species the fibre tracts have in general a high benzoylcholine-hydrolysing activity.

It is of interest in this connection to note that Koelle (1951), using a histochemical method, has reported the presence of pseudo-cholinesterase in the glial cells of sensory and autonomic ganglia and also in the Schwann cells of myelinated nerves. In confirmation of these results Thompson and Webster (1952) have recently examined a number of gliomas by the standard Warburg technique; each of these tumours was found to be rich in pseudo-cholinesterase. The only oligodendroglioma in the series contained a higher activity (2845 µl.CO_2/g./hr., with butyrylcholine as substrate) than any other area of the human nervous system so far studied, while giving a value of only 120 µl.CO_2/g./hr. for true cholinesterase (acetyl-β-methylcholine as substrate).

Boell and Nachmansohn (1940) have studied the localization of cholinesterase in the nerve fibre, using the giant fibres of the squid, and have found it to be present almost entirely in the sheath, the extruded axoplasm showing a very low level of activity.

From this work on the distribution of cholinesterase types two facts stand out clearly:

(1) that pseudo-cholinesterase is very widely distributed throughout the body, including both the central and peripheral nervous systems;

(2) that the true cholinesterase, although mainly restricted to nerve and muscle, is not confined to these tissues. Not only do the red blood cells of many species contain it, but also a number of other tissues. It might be thought that the relatively low level of true cholinesterase in these other tissues is evidence of the association of this enzyme with the few nerve fibres present in these tissues. It must be remembered, however, that in addition to the erythrocytes, perfused blood-free placenta, a tissue that is usually regarded as being nerve-free, contains mainly the true enzyme (Ord and Thompson, 1950b).

resulting in demyelination, it would be expected that poisoning with other selective inhibitors of this enzyme should also lead to loss of myelin and paralysis, and this has now been shown to be the case for two other anti-cholinesterases. Bidstrup and Hunter [illegible] about two weeks after exposure to the poison, which was said to resemble clinically that following poisoning by TOCP. Further, Barnes and Denz (1953) have succeeded in producing demyelination and paralysis in hens both by this compound and by DFP; they found it necessary to atropinize their hens in order to prevent undue disturbance from acetylcholine accumulation, and so allow the birds to survive and show the late development of paralysis. Apart from a report of the production of late paralysis in two dogs following the repeated administration of DFP (Koelle and Gilman, 1946a) the development of these neurological changes in animals poisoned with DFP does not seem to have been previously described. This may have been because the poisoned animals died from the effects of acetylcholine accumulation in the early stages of the intoxication, before sufficient time had elapsed for these late effects to appear, while in the case of the survivors the dose of the poison may have been too small to allow them to develop.

Poisoning with certain other inhibitors of pseudo-cholinesterase has, however, failed to produce demyelination (Barnes and Denz, 1952; Thompson and Webster, 1952), so that if inactivation of this enzyme is playing any causative role in the development of the demyelination it would seem that it cannot be the only factor. It is possible that TOCP and these other demyelinating compounds may have some other action, and that pseudo-cholinesterase inhibition is merely incidental. But, as regards TOCP, inhibition of pseudo-cholinesterase and tributyrinase (Hottinger and Bloch, 1943) are the only biochemical effects which have so far been described. Further, Earl, Thompson and Webster (1953) have shown that brain tributyrinase is very much less sensitive to inhibition by TOCP than is pseudo-cholinesterase, while glucose and pyruvate

they found that concentrations of TOCP, which caused almost complete inhibition of the pseudo-cholinesterase in the cerebrum, spinal cord, sciatic nerve and serum, produced only a negligible inhibition of the true cholinesterase of nerve tissue, striated muscle or erythrocytes (Table 3).

TABLE 3. Inhibition of Human Cholinesterases by Tri-*ortho*-cresyl Phosphate

	% Inhibition produced by			
	50	*100*	*200*	*500*
	µg. TOCP/3ml.			
Pseudo-cholinesterase				
Serum	55	73	93	—
Cerebrum	65	81	91	—
Spinal cord	55	64	74	—
Sciatic nerve	81	92	99	—
True cholinesterase				
Erythrocytes	7	4	7	4
Cerebrum	4	9	8	—
Spinal cord	—	2	7	2
Striated muscle	3	8	10	11

The levels of true and pseudo-cholinesterase of the brain and spinal cord of hens killed at varying intervals after poisoning with TOCP were next compared with those of normal birds (Earl and Thompson, 1952b); the pseudo-cholinesterase activity was found to be markedly lowered as early as one day after poisoning and to remain at a low level for at least ten days; but again, under these *in vivo* conditions, the true cholinesterase in these tissues was hardly affected.

These results do not therefore support Bloch's explanation of the mechanism underlying the paralysis, inasmuch as muscle cholinesterase does not seem to be seriously affected by TOCP. But since, as already pointed out, pseudo-cholinesterase is predominantly associated with the white matter of the C.N.S., and since in human peripheral nerves it is present in significantly greater amounts than the true (Thompson and Webster, 1952), it is possible that it may be concerned with some aspect of metabolism connected with myelin formation and turnover.

If inhibition of this enzyme in nerve tissue is to be regarded as

that dogs remain symptom-free even when the pseudo-cholinesterase in the pancreas and the superior cervical ganglia is inhibited to the extent of 80–90 per cent. From these observations it has been concluded by some workers that pseudo-cholinesterase does not play any important part in the destruction of acetylcholine released during the normal transmission processes inside the body.

The effects produced by nerve section have also been interpreted as minimizing the significance of pseudo-cholinesterase in synaptic transmission. Thus, Sawyer and Hollinshead (1945) have reported that pre-ganglionic nerve section results in a profound fall in the level of true cholinesterase in the superior cervical ganglion of the cat, while causing relatively little effect on the level of the pseudo-cholinesterase. Sawyer (1946) has also shown that in the Wallerian degeneration that follows section of the guinea-pig sciatic nerve, pseudo-cholinesterase activity is unaffected although a substantial reduction of the true enzyme occurs. On the other hand, it has recently been found that when the nerves to the rabbit's ear are sectioned there is a marked fall in the pseudo-cholinesterase activity of the ear arteries, resulting in a complete or almost complete absence of activity by about 28 days after the nerve section (Armin, Grant, Thompson and Tickner, 1953).

More decisive evidence suggesting a relation between pseudo-cholinesterase activity and cell function has been obtained by Koelle, Koelle and Friedenwald (1950); using the cat's ileum they have shown that concentrations of DFP too low to cause any appreciable inhibition of the true cholinesterase but capable of reducing the pseudo-cholinesterase activity substantially cause an increase in tone and also increased amplitude of the contractions of the isolated organ; from these observations they have suggested that pseudo-cholinesterase may act by supplementing the action of the true enzyme in catalysing the hydrolysis of endogenously liberated acetylcholine.

In this connection it is of interest to recall that Feldberg and Lin (1950) have claimed that the spontaneous release of acetylcholine by the intestine is an example of acetylcholine metabolism of largely non-nervous origin, and have demonstrated

oxidation by brain, amine oxidase, trypsin, lecithinase and cephalinase are *all* unaffected even by high concentrations of TOCP. Hottinger and Bloch (1943) had earlier shown that serum alkaline phosphatase and pancreatic lipase are also insensitive.

PHYSIOLOGICAL FUNCTIONS OF THE CHOLINESTERASES

The evidence, obtained both from distribution studies in the nervous system and in muscle and from observations on the effects of anti-cholinesterases, is strongly in favour of the view that the true cholinesterase is concerned with the inactivation of physiologically released acetylcholine at motor end-plates and at synapses, both peripheral and central. When the activity of this enzyme is prevented, signs attributable to acetylcholine accumulation in the tissues develop. The true cholinesterase of the central nervous system must, however, be inhibited to a very considerable extent before any effect can be observed in the living animal (Mazur and Bodansky, 1946; Koelle and Gilman, 1946b; Nachmansohn and Feld, 1947). Since more than 80 per cent inhibition must be brought about before signs are apparent, it would seem that there is normally present an ample margin of cholinesterase activity in the central nervous system.

Nachmansohn and his colleagues (see Nachmansohn and Wilson, 1951) have put forward the theory that acetylcholine is involved not only in synaptic transmission but also in the conduction of the nerve impulse along the nerve fibre. True cholinesterase is certainly present in the nerve as well as at nerve endings, and they have suggested that its function in the nerve fibre is essentially the same as that at the nerve-ending. This theory has, however, been criticized, and it is not possible here to review the arguments for and against it.

The position with regard to the part played by the pseudocholinesterases in physiological processes is obscure. It has been repeatedly shown both in animals and in man that the serum cholinesterase can be almost completely inactivated by DFP without the development of any significant symptoms (Mazur and Bodansky, 1946). Hawkins and Gunter (1946) have found the same absence of signs of acetylcholine accumulation after inhibition of the serum enzyme by Nu683; they have also stated

FELDBERG, W. S. (1950). *Proc. roy. Soc.* B. 137, 285.
FELDBERG, W. S. and LIN, R. C. Y. (1950). *J. Physiol.* 111, 96.
GLICK, D. (1941). *Nature, Lond.* 148, 662.
HAWKINS, R. D. and GUNTER, J. M. (1946). *Biochem. J.* 40, 192.
HAWKINS, R. D. and MENDEL, B (1949). *Biochem J.* 44, 260.
HOTTINGER, A. and BLOCH, H. (1943). *Helv. Chim. Acta*, 26, 142
KOELLE, G. B. (1950). *J. Pharm. Exp Ther.* 100, 158
KOELLE, G. B. (1951). *J. Pharm. Exp Ther.* 103, 153.
KOELLE, G. B. and GILMAN, A. (1946a) *J. Pharm Exp. Ther.* 87, 435
KOELLE, G. B. and GILMAN, A. (1946b). *J. Pharm. Exp Ther* 87, 421.
KOELLE, G. B., KOELLE, E S. and FRIEDENWALD, J. S (1950). *J. Pharm. Exp. Ther.* 100, 180.
LANGEMANN, K. (1944). *Helv. Physiol. Pharmacol. Acta*, 2, C17.
LEVINE, M. G. and SURAN, A. A. (1950). *Nature, Lond.* 166, 698
MAZUR, A. and BODANSKY, O. (1946) *J biol. Chem* 163, 261
MENDEL, B. and MUNDELL, D. B. (1943). *Biochem. J.* 37, 64.
MENDEL, B., MUNDELL, D. B. and RUDNEY, H. (1943). *Biochem. J.* 37, 473.
MENDEL, B. and RUDNEY, H (1943a). *Biochem. J.* 37, 59
MENDEL, B. and RUDNEY, H. (1943b). *Science*, 98, 201.
MENDEL, B and RUDNEY, H. (1944). *Science*, 100, 499.
NACHMANSOHN, D. (1939). *Bull. soc. chim. biol. Paris*, 21, 761.
NACHMANSOHN, D and FELD, E A (1947) *J. biol Chem.* 171, 715.
NACHMANSOHN, D and ROTHENBERG, M A. (1944) *Science*, 100, 454
NACHMANSOHN, D. and ROTHENBERG, M A. (1945). *J. biol. Chem* 158, 653
NACHMANSOHN, D. and WILSON, I B (1951) *Advances in Enzymology*, 12, 259, Interscience New York.
ORD, M. G. and THOMPSON, R H. S. (1950a). *Biochem J* 46, 346.
ORD, M G and THOMPSON, R. H S (1950b). *Nature, Lond* 165, 927
ORD, M. G and THOMPSON, R H S. (1951) *Biochem J* 49, 191.
ORD, M G and THOMPSON, R H S. (1952). *Biochem J* 51, 245.
RICHTER, D. and CROFT, P. G (1942). *Biochem. J.* 36, 746.
SAWYER, C. H. (1946) *Amer. J Physiol.* 146, 246.
SAWYER, C H. and EVERETT, J W (1947) *Amer J. Physiol.* 148, 675.
SAWYER, C. H. and HOLLINSHEAD, H (1945). *J Neurophysiol.* 8, 137
SMITH, M I and ELVOVE, E (1930). *U.S Publ Hlth Rep* 43, 1703
SMITH, M I, ELVOVE, E and FRAZIER, W. H (1930) *U.S Publ Hlth. Rep* 43, 2509.
SMITH, M I. and LILLIE, R D (1931). *Arch. Neurol. Psychiat, Chicago*, 26, 976.
THOMPSON, R H. S (1947) *J Physiol.* 105, 370.
THOMPSON, R. H S. and TICKNER, A (1953) *J. Physiol* 121, 623
THOMPSON, R. H S and WEBSTER, G R (1952) Unpublished observations.
THOMPSON, R H S. and WHITTAKER, V P (1944) *Biochem J* 38, 295
WHITTAKER, V. P. (1951) *Physiol. Rev* 31, 312
ZELLER, E A (1949). *Helv chim. Acta*, 32, 484
ZELLER, E A. and BISSEGGER, A (1943) *Helv. chim. Acta*, 26, 619

that the choline acetylase activity of this organ is not dependent on the presence of nerve fibres, the highest activity in any part of the wall being present in glandularis mucosae. Ord and Thompson (1950a) found that the level of pseudo-cholinesterase in intestine muscle and mucosa is very high, and Feldberg (1950) has suggested that this largely non-nervous acetylcholine metabolism of the intestine may be responsible for stimulating both the spontaneous contractions of the gut wall and the secretion of the succus entericus. Although we have no proof as yet, it seems possible that such a 'local hormone' conception (see Burn, 1950) of the role of acetylcholine might represent a widely distributed non-nervous mechanism for the stimulation of cells, and that pseudo-cholinesterase might in certain situations be the hydrolysing enzyme concerned with this local function.

REFERENCES

ADRIAN, E. D., FELDBERG, W. and KILBY, B. A. (1947). *Brit. J. Pharmacol.* 2, 56.

ALLES, G. A and HAWES, R. C. (1940). *J. biol. Chem.* 133, 375.

ARMIN, J., GRANT, R. T., THOMPSON, R. H. S. and TICKNER, A. (1953). *J. Physiol.* 121, 603.

AUGUSTINSSON, K. B (1948). *Acta. Physiol. Scand.* 15, Suppl. 52.

BARNES, J M. and DENZ, F. A. (1952). Unpublished observations.

BARNES, J M and DENZ, F A (1953). *J. Path. Bact.* 65, 597.

BIDSTRUP, P. L and HUNTER, D. (1952). *Lancet*, 1, 262.

BLOCH, H. (1941). *Helv. Med. Acta*, 8, Suppl. 7, 15

BODANSKY, O. (1946). *Ann N.Y. Acad. Sci.* 47, 521.

BOELL, E. J. (1945). *J. cell. comp. Physiol.* 25, 75.

BOELL, E. J. and NACHMANSOHN, D. (1940). *Science*, 92, 513.

BURGEN, A. S. V and CHIPMAN, L. M. (1951). *J. Physiol.* 114, 296.

BURN, J. H. (1950). *Proc. roy Soc.* B. 137, 281

EARL, C. J. and THOMPSON, R H. S. (1952a). *Brit. J. Pharmacol.* 7, 261.

EARL, C. J. and THOMPSON, R. H. S. (1952b). *Brit. J. Pharmacol.* 7, 685.

EARL, C. J., THOMPSON, R. H S. and WEBSTER, G. R (1953). *Brit. J. Pharmacol.* 8, 110.

FELDBERG, W. S. (1950). *Proc. roy. Soc.* B. 137, 285.
FELDBERG, W. S. and LIN, R. C. Y. (1950). *J. Physiol.* 111, 96.
GLICK, D. (1941) *Nature, Lond.* 148, 662
HAWKINS, R. D. and GUNTER, J. M. (1946). *Biochem J.* 40, 192.
HAWKINS, R. D. and MENDEL, B (1949). *Biochem J.* 44, 260.
HOTTINGER, A. and BLOCH, H. (1943). *Helv. Chim Acta*, 26, 142.
KOELLE, G. B. (1950). *J. Pharm. Exp Ther* 100, 158.
KOELLE, G. B. (1951). *J. Pharm. Exp. Ther.* 103, 153.
KOELLE, G. B. and GILMAN, A. (1946a) *J. Pharm. Exp. Ther.* 87, 435.
KOELLE, G. B and GILMAN, A. (1946b). *J Pharm. Exp Ther.* 87, 421.
KOELLE, G. B., KOELLE, E. S. and FRIEDENWALD, J. S (1950). *J. Pharm. Exp. Ther* 100, 180.
LANGEMANN, K. (1944) *Helv. Physiol. Pharmacol. Acta*, 2, C17.
LEVINE, M. G. and SURAN, A. A (1950). *Nature, Lond.* 166, 698.
MAZUR, A. and BODANSKY, O. (1946). *J. biol. Chem.* 163, 261
MENDEL, B. and MUNDELL, D. B. (1943) *Biochem J.* 37, 64.
MENDEL, B , MUNDELL, D B. and RUDNEY, H (1943). *Biochem. J.* 37, 473.
MENDEL, B and RUDNEY, H. (1943a). *Biochem. J* 37, 59.
MENDEL, B. and RUDNEY, H. (1943b) *Science*, 98, 201.
MENDEL, B. and RUDNEY, H. (1944). *Science*, 100, 499.
NACHMANSOHN, D. (1939). *Bull. soc chim. biol Paris*, 21, 761.
NACHMANSOHN, D. and FELD, E A (1947). *J. biol. Chem.* 171, 715
NACHMANSOHN, D and ROTHENBERG, M A (1944) *Science*, 100, 454.
NACHMANSOHN, D and ROTHENBERG, M. A (1945) *J biol. Chem* 158, 653
NACHMANSOHN, D and WILSON, I. B (1951) *Advances in Enzymology*, 12, 259, Interscience. New York
ORD, M. G. and THOMPSON, R H S. (1950a) *Biochem. J.* 46, 346
ORD, M G and THOMPSON, R H S. (1950b) *Nature, Lond* 165, 927.
ORD, M. G and THOMPSON, R H. S. (1951) *Biochem J* 49, 191.
ORD, M G and THOMPSON, R H S (1952) *Biochem J* 51, 245
RICHTER, D. and CROFT, P. G (1942). *Biochem J* 36, 746
SAWYER, C. H. (1946). *Amer. J Physiol* 146, 246.
SAWYER, C H and EVERETT, J. W (1947) *Amer. J. Physiol.* 148, 675.
SAWYER, C H. and HOLLINSHEAD, H. (1945) *J Neurophysiol* 8, 137
SMITH, M I. and ELVOVE, E (1930) *U S Publ Hlth Rep* 43, 1703.
SMITH, M. I., ELVOVE, E. and FRAZIER, W H (1930). *U S Publ Hlth Rep.* 43, 2509
SMITH, M I. and LILLIE, R D. (1931) *Arch. Neurol. Psychiat , Chicago*, 26, 976.
THOMPSON, R H S (1947). *J Physiol* 105, 370
THOMPSON, R H S. and TICKNER, A (1953) *J Physiol* 121, 623
THOMPSON, R. H. S and WEBSTER, G. R (1952). Unpublished observations
THOMPSON, R. H S and WHITTAKER, V P (1944) *Biochem J* 38, 295
WHITTAKER, V P (1951) *Physiol. Rev* 31, 312.
ZELLER, E A. (1949). *Helv chim. Acta*, 32, 484
ZELLER, E. A. and BISSEGGER, A. (1943) *Helv. chim Acta*, 26, 619.

X

The Metabolism of the Adrenocortical Hormones

G. F. MARRIAN

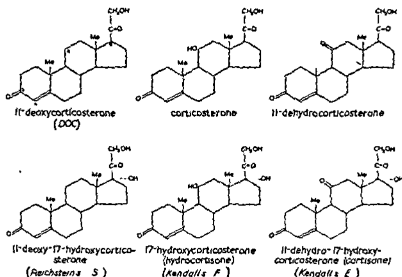
Introduction

SINCE the discovery by Hench, Kendall and their co-workers (1949) of the remarkable effects of cortisone in the treatment of rheumatoid arthritis and certain related diseases there has been such a spate of lectures on the adrenocortical hormones that I almost feel that another lecture on this topic should be prefaced by an apology! The only excuse I can offer for what I am about to do is that the particular aspect of this topic which I am going to talk about—the biochemistry of the adrenocortical hormones—is one which I still find to be absorbingly interesting myself—in spite of all the lectures I have heard during the past few years.

I must begin by reminding you briefly of a few facts concerning the physiological roles of the adrenocortical hormones and their chemical nature.

Bilateral adrenalectomy is usually fatal in experimental animals within one to two weeks, but before death certain characteristic metabolic abnormalities may be observed. On the one hand there are marked disturbances in the urinary excretion of certain electrolytes. The excretion of Na^+ and Cl^- is greatly increased, and this results in a decreased concentration of these ions in the blood plasma and extracellular fluids, and consequently there is a decrease in the volumes of both the latter. The excretion of K^+ is, in contrast, diminished; and this leads to an increase in the K^+ content of the plasma and extracellular

fluids, and secondarily an increase in K^+ in the cells. On the other hand there are marked disturbances in carbohydrate metabolism. The capacity of the liver to store glycogen is impaired, the rate of gluconeogenesis from protein is decreased, and also, perhaps, the rate of carbohydrate utilization in the tissues is increased.



Biologically Active Steroids Isolated from the Adrenal Cortex

The lives of adrenalectomized animals may be prolonged and these abnormalities in salt and carbohydrate metabolism corrected by the daily administration of suitably prepared extracts of adrenocortical tissue. If excessive amounts of such extracts are administered, even to normal animals, metabolic abnormalities of the opposite nature may be induced—i.e. decreased excretion of Na^+ and Cl^-, and increased excretion of K^+, increased gluconeogenesis from protein, and increased glycogen storage in the liver.

Six different pure steroids have been isolated from adrenal extracts each of which is active in respect to salt excretion and/or carbohydrate metabolism.

Of these DOC is by far the most active in its effect upon salt

excretion, while hydrocortisone and cortisone have the greatest effect upon carbohydrate metabolism. The two types of physiological activity displayed by these steroids—towards salt excretion on the one hand, and towards carbohydrate metabolism on the other, can conveniently be called 'mineralocorticoid' and 'glucocorticoid' activities respectively.

This is not the whole story of the active compounds in the adrenal cortex. Many years ago it was shown that a high proportion of the total activity of a crude adrenal extract remains in the amorphous fraction after as complete removal as possible of the crystalline steroids. This unknown active factor in the amorphous fraction was shown to be more potent as a mineralocorticoid than as a glucocorticoid, and to be much less stable than any of the six active crystalline steroids. Because of its instability, perhaps, it stubbornly resisted all attempts made to isolate it in a pure crystalline form. Recently, however, very important preliminary work has been carried out in two laboratories in London which suggests that we will soon know what the active material in the amorphous fraction is composed of. At the National Institute for Medical Research, Bush (1952) has examined preparations of the amorphous fraction by paper chromatographic methods and has detected the presence of three substances in it which are probably steroids and which are probably more highly oxygenated than even cortisone or hydrocortisone. One of these has been tested for glucocorticoid activity using the cold-exposure test of Selye and Schenker (1938), and has been found to be very potent. [illegible]
carried out at the Middles[illegible]
their co-workers (1952). [illegible]
methods they have detected a new steroid in adrenal extracts which is twenty-five times more active as a mineralocorticoid even than DOC, and like the unknown active principle in the amorphous fraction it is rather unstable. Curiously enough it does not appear to contain the αβ unsaturated ketonic grouping possessed by the other active adrenocortical steroids.[1] Whether

[1] Later work by these authors indicates that this new compound, now known as *electrocortin*, may after all be a Δ^4-3-ketone, and may be even more active as a mineralocorticoid than was originally believed.

this new mineralocorticoid is identical with Bush's new glucocorticoid is not clear yet.

ADRENOCORTICAL HORMONES IN BLOOD

The fact that a number of different physiologically active steroids have been isolated from extracts of the adrenal gland does not by itself justify us in labelling them 'adrenocortical hormones'. A hormone is a physiologically active substance secreted by an endocrine gland into the blood, and accordingly we should only apply the term 'hormone' to an active adrenocortical steroid if it has been isolated from or detected in blood.

The question of which of the various active adrenocortical steroids can strictly be regarded as hormones is not one which is merely of interest to those concerned with terminology. On the contrary it is of very real physiological importance in connection with the speculative but most interesting theories developed in recent years by Selye of Montreal on the role of the adrenal cortex in the aetiology of the so-called 'diseases of adaptation'. Selye (1949a, b) has suggested that an important aetiological factor in such diseases is an abnormal adaptive response of the adrenal cortex to prolonged stress, resulting in an alteration in the normal balance between mineralo- and glucocorticoids in the secretion of the glands. In 1949 when this theory was advanced in its completed form there was no evidence to suggest that diseases of adaptation were preceded or accompanied by any alteration in the make-up of the adrenocortical secretion; in fact at that time nothing was known concerning the nature of the active adrenocortical steroids secreted by the gland into the blood. Necessarily, therefore, this interesting theory had to be judged, not as an actual contribution to knowledge, but rather on its merits as a stimulus to further research.

During the past two years the problem of the chemical nature of the active adrenocortical steroids in blood has been intensively studied in several different laboratories with conspicuous success. Nelson, Samuels and their co-workers (1950, 1951) at Salt Lake City, using orthodox chemical isolation methods and paper chromatography, have shown with certainty the presence of corticosterone and hydrocortisone in the adrenal-vein blood of

dogs. They also believe they have detected hydrocortisone in the peripheral blood of human subjects. Bush (1952) has also detected corticosterone and hydrocortisone in the adrenal-vein blood of dogs, and also the same new glucocorticoid which he had detected in gland extracts; while Grundy, Simpson and their co-workers (1952), also working with blood from the adrenal vein of dogs, have detected their new highly active mineralocorticoid.

All this is an excellent beginning, but I think it is no more than a beginning. It is not improbable, in my opinion, that when better methods have been developed for extracting blood and for fractionating the extracts, the presence of other active adrenocortical steroids may be detected; and it may be possible to elaborate methods for the quantitative determination of these steroids in the peripheral blood of human subjects so that Selye's theory can be put to direct experimental test.

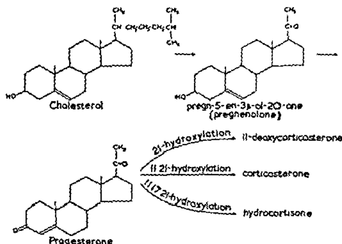
The Biogenesis of the Adrenocortical Hormones

The adrenal cortex does not contain any very considerable reserve stock of its own hormones, but its capacity for synthesizing these hormones rapidly is remarkable. Thus Vogt (1947, 1951), in experiments in which she determined the glucocorticoid content of the adrenal-vein blood of dogs, has shown that the output of the gland per minute is nearly ten times greater than the amount of glucocorticoid present in the gland itself at any one time.

The precursor of the adrenocortical hormones which is stored in the glands and which can be rapidly converted into these hormones is cholesterol, and the evidence leading to this conclusion is briefly as follows.

Sayers, Long and their co-workers (1944, 1946) showed nearly ten years ago that the administration to experimental animals of the adrenocorticotrophic hormone of the anterior lobe of the pituitary (ACTH) is followed by a marked fall in the amount of cholesterol in the adrenal glands, and a marked increase in the amount of glycogen stored in the liver. It was reasonable to assume from these findings that on stimulation by ACTH the adrenal glands had manufactured glucocorticoids from cholesterol in increased amounts and secreted them.

More direct evidence for the formation of adrenocortical hormones from cholesterol has been obtained in recent years by Hechter, Pincus and their co-workers (1951, 1952) at the Worcester Foundation in Massachusetts. These workers perfused isolated beef adrenal glands with blood to which had been added cholesterol tagged at position 3 with radioactive ^{14}C.



Suggested Route for the Formation of the Adrenocortical Hormones from Cholesterol

Using chromatographic methods they were able to show the presence in the adrenal effluent of corticosterone and hydrocortisone both of which were found to contain significant amounts of radioactivity. Their experiments would seem to provide clear-cut proof that cholesterol can be degraded in the glands to the adrenocortical hormones, but they do not exclude the possibility that like cholesterol itself these hormones may also be synthesized in the glands directly from 2-carbon fragments.

The same group of workers have also been carrying out experiments perfusing a variety of different steroids through isolated adrenal glands in order to gain some insight into possible intermediate stages in the degradation of cholesterol to the hormones.

Their findings are compatible with the idea that the side-chain of cholesterol becomes oxidized away to give pregnenolone; that the latter is then further oxidized to progesterone, which then undergoes hydroxylation at position 21 to give DOC, at positions 21, 17 and 11 to give hydrocortisone, or at positions 21 and 11 to give corticosterone. The hormones with ketonic groups at C-11 were also found in their perfusates in much smaller amounts than the corresponding 11-hydroxy compounds, and accordingly they suggest that the latter are the primary products while the 11-keto compounds are secondary ones formed by the oxidation of these.

By carrying out perfusion experiments both with and without added ACTH the Worcester Foundation group have attempted to find out at which stage in the series of reactions ACTH acts. Their present view is that it stimulates the degradation of cholesterol to pregnenolone or progesterone, and probably plays no part in the subsequent hydroxylations. There is, however, a serious difficulty about accepting this view unreservedly as they themselves realize. There is good reason to believe that progesterone is formed in the corpus luteum from cholesterol and that pregnenolone is an intermediary in this process. If ACTH stimulates the formation of pregnenolone and progesterone from cholesterol in the adrenal, why does it not also act like luteotrophin in stimulating progesterone secretion by the corpus luteum? This is not an unsuperable difficulty, but it seems to be a very real one.

During the past two years much has been learned about the enzyme systems in the adrenal cortex responsible for the hydroxylation at C-11. This system appears to be located in the mitochondria (Sweat, 1951) and washed preparations of adrenocortical mitochondria will effect 11-hydroxylations of various steroids with considerable efficiency in the presence of the right co-factors. Mg ions are necessary for the system and also fumarate or some other component of the tricarboxylic acid cycle such as malate. Apparently the 11-hydroxylation system is coupled in some way with the tricarboxylic acid cycle.

One of the most interesting of all the biochemical problems in connection with the adrenal cortex—an unsolved problem, is

that of the role of ascorbic acid in the biogenesis of the adrenocortical hormones.

It has been known for over twenty years that large amounts of ascorbic acid are concentrated in the adrenal cortex, and Sayers, Long and their co-workers (1944, 1946) in their work on the effect of ACTH on the adrenal cholesterol observed that the fall in adrenal cholesterol was preceded by a marked fall in the adrenal ascorbic acid. The effect of ACTH in causing a rapid fall in adrenal ascorbic acid has, of course, been amply confirmed since, and it forms the basis for the well-known Sayers method for the standardization of ACTH preparations.

Since ACTH administration causes a fall in adrenal ascorbic acid and cholesterol and an increased secretion of adrenocortical hormones, it was thought at first by many that ascorbic acid might be a necessary part of the oxidation system which converts cholesterol to the hormones. The evidence, most of which is indirect, on whether ascorbic acid is necessary for the biogenesis of the adrenocortical hormones is conflicting, however. All this evidence cannot be discussed in detail now, but some of it must be, since it has recently given rise to some very interesting speculation.

Before I go on to discuss this evidence I must digress at some length to say something about the steroids of adrenocortical origin which are excreted in the urine. These are of various types and they may be classified as follows according to the kind of side chain which they have at C-17.

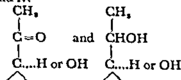
i.	CH_2OH | C=O | C .. H or OH	Unchanged adrenocortical hormones—cortisone, hydrocortisone and perhaps others. Also partial reduction products of these in which the Δ3 unsaturated ketone grouping in ring A has been fully or partly reduced.
ii.	CH_2OH | CHOH | C . H or OH	As above, but with the C-20 ketone group reduced

Steroids of these types on oxidation with periodic acid yield formaldehyde from the terminal $—CH_2OH$ of the side chain, and their amounts in urinary extracts can be roughly determined by

oxidation of extracts with periodic acid and colorimetric determination of the formaldehyde evolved.

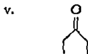
The methods for the quantitative determination of such 'formaldehydogenic' steroids in urine are very far from satisfactory, but bad though they are, there is ample evidence that they provide a rough means of assessing the secretory activity of the adrenal glands.

iii. and iv.



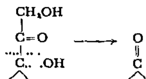
Formed from the adrenocortical hormones by complete reduction at C-21. Also perhaps secreted as such by the adrenals.

v.



The 17-ketosteroids. In men the urinary 17-ketosteroids are partly derived from the adrenal cortex and partly from the testes. In women they are wholly derived from the adrenal cortex.

The adrenocortical 17-ketosteroids probably originate by the metabolic oxidation of the side chain of the 17-hydroxycorticosteroids



and may also be secreted as such by the glands.

The methods for the quantitative determination of the 17-ketosteroids in urine are fairly reliable and they provide a good means of assessing the secretory activity of the adrenal gland.

We can now return to consider the problem of the *rôle* of ascorbic acid in the biogenesis of the adrenocortical steroids.

Using guinea-pigs, which are the only animals other than primates which cannot synthesize their own ascorbic acid, Nadel and Schneider (1952a, b) have recently attempted to find out whether ascorbic acid is or is not necessary for the biogenesis of the adrenocortical steroids, by following the daily excretion of formaldehydogenic steroids while the animals were fed on an ascorbic acid-free diet. Their findings were surprising, because after about 20–25 days on the diet, when the animals had

developed marked symptoms of scurvy, the daily excretion of formaldehydogenic substances had significantly increased. This increase was a fivefold one, and therefore although the method for determining formaldehydogenic steroids is admittedly bad, there is no doubt that this increase is real. Furthermore Nadel and Schneider showed that the additional formaldehydogenic substances excreted had the properties of hydrocortisone.

At about the same time Clayton and Prunty (1951), also using guinea-pigs, showed that the 17-ketosteroid excretion greatly increased after prolonged feeding on an ascorbic acid-free diet.

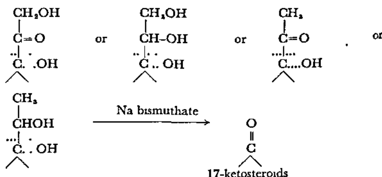
There are now, therefore, two independent pieces of fairly direct evidence indicating that when the adrenal glands are depleted of ascorbic acid their secretory activity is not, as one might have anticipated, decreased, but on the contrary very markedly increased. It is relevant here to recall that it has been known for many years that in scurvy the adrenal cholesterol is markedly diminished in amount although the glands themselves enlarge; and accordingly it might almost be supposed that the role of ascorbic acid in the adrenal cortex is not to aid in the formation of the adrenocortical hormones from cholesterol, but rather to put a brake on this process in order to prevent it proceeding too rapidly. This interesting speculation was recently put forward by a colleague of mine at Edinburgh, Dr. Peter Meiklejohn, and I would like to carry it a stage further by suggesting that conceivably the role of ACTH might be to combine with or inactivate in some way the adrenal ascorbic acid, thus allowing the breakdown of cholesterol to adrenocortical hormones to proceed more rapidly.

Recent Progress in the Quantitative Determination of the Urinary Adrenocortical Steroids

In the whole of clinical chemistry there is no problem more urgently in need of solution than that of devising really satisfactory methods for the quantitative determination in urine of adrenocortical steroids other than the 17-ketosteroids. Time will not permit me to discuss this problem in any detail, but I want to conclude by referring briefly to two recent contributions in this field which I believe may be important.

The greater part of the adrenocortical steroids in urine are present not in the free state, but conjugated with glucuronic acid, and perhaps also with sulphuric acid. As a necessary preliminary to the extraction of these steroids from urine and their subsequent determination the conjugates have to be hydrolysed; and this is where the chief difficulty lies, since it is known that a considerable fraction of the biologically active and formaldehydogenic steroids in urine becomes destroyed if the usual procedure for the hydrolysis of steroid conjugates—boiling the urine with acid—is employed. Many workers have in the past attempted to avoid this destruction by hydrolysing the conjugates with acid at room temperature. This is a most inefficient procedure, however. A better method of hydrolysing these conjugates is by incubating the urine with preparations of the enzyme β-glucuronidase—an enzyme specific for the hydrolysis of β-glucuronides. The optimum conditions for the enzymic hydrolysis have not yet been fully worked out, however, and in any case it is doubtful if this method could be readily adapted to the needs of the routine clinical chemical laboratory.

Brooks and Norymberski (1952) have recently found that all C_{21} 17-hydroxyadrenocortical steroids yield 17-ketosteroids on oxidation with sodium bismuthate



and they have successfully applied this reaction to the determination of such steroids in urine. The oxidation is carried out on the untreated urine, and the 17-ketosteroid conjugates thus formed are then hydrolysed in the usual way by boiling with acid. In this way they avoid the difficulty of destruction of the C_{21}

steroids by hot acid. On another sample of urine an ordinary determination of the 17-ketosteroids is carried out, and the difference in the two values gives them a figure for the 17-hydroxysteroids—i.e. what they term the '17-ketogenic steroids'. This is an ingenious way around the hydrolysis difficulty, and the method should be a useful one for routine purposes.

Another useful method has recently been devised by Tompsett. He has found that whereas cortisone is rapidly destroyed by hot acid, DOC is comparatively stable; and on the basis of this finding has worked out a method employing hot-acid hydrolysis for the determination in urine of the acid-stable 'DOC-like' formaldehydogenic steroids. This method has yielded interesting results when used as an experimental tool by Stewart, Robson and Tompsett (1952). Applying the procedure to the urine of human subjects they have recorded normal values of 'DOC-like' substances of about 5 mg. per 24 hours, and they have found the values to be very greatly increased when the subjects were put on a salt-free diet. Evidently the 'DOC-like' substances they are determining are steroids specifically related to salt metabolism, and this work provides fresh evidence that the adrenal cortex does secrete specific mineralocorticoids.

Unpublished work carried out recently by Mrs. Eileen Michie and myself on a whole series of different C_{21} adrenocortical steroids, including cortisone, hydrocortisone, DOC, and Reichstein's 'S', indicates quite clearly that instability to hot acid depends upon the presence of a 17-hydroxyl group: the 17-deoxy compounds appear to be reasonably stable to hot acid. It may be assumed therefore that Tompsett's method specifically determines steroids which like DOC are 17-deoxy ones.

These two new methods, in which the difficulty over the hydrolysis of conjugates in urine is not overcome, but dodged, have provided us with simple procedures for the determination in urine of all C_{21} 17-hydroxysteroids on the one hand, and for the formaldehydogenic 17-deoxy steroids on the other.[1]

[1] Since writing the above the present authors (Marrian and Atherden, 1953) have provided evidence to indicate that the hot acid-stable formaldehydogenic substances of Tompsett are artefacts formed from 17-hydroxysteroids rather than 17-deoxysteroids as the latter believed

REFERENCES

BROOKS, C. J. W. and NORYMBERSKI, J. K. (1952). *Chem. & Ind.* 80.
BUSH, J. E. (1952). *Biochem. J.* 50, 370.
CLAYTON, B. E. and PRUNTY, F. T. G. (1951). *Brit. med. J.* ii, 927.
GRUNDY, H. M. and SIMPSON, S. A. (1952). *Nature*, 169, 795.
GRUNDY, H. M., SIMPSON, S. A., TAIT, J. F. and WOODFORD, M. (1952). *Acta Endocrinol.* 11, 199.
HECHTER, O., ZAFFARONI, A., JACOBSEN, R. P., LEVY, H., JEANLOZ, R. W., SCHENKER, V. and PINCUS, G. (1951). *Rec. Prog. Horm. Res.* 6, 215.
HECHTER, O. M. (1952). 'Adrenal Cortex', Trans. Third Conf. of Josiah Macy Foundation, New York, p. 115.
HENCH, P. S., KENDALL, E. C., SLOCUMB, N. H. and POLLEY, H. F. (1949). *Proc. Mayo Clin.* 24, 181.
MARRIAN, G. F. and ATHENDEN, S. M. (1953). *Acta Endocrinol.* 13, 265.
MEIKLEJOHN, A. P. Personal communication.
NADEL, E. M. and SCHNEIDER, J. J. (1952a). *Fed. Proc.* 11, 263.
NADEL, E. M. and SCHNEIDER, J. J. (1952b). *Endocrinol.* 51, 5.
NELSON, D. H., REICH, H. and SAMUELS, L. T. (1950). *Science*, 111, 578
NELSON, D. H, SAMUELS, L. T., WILLARDSON, D. and TYLER, F. H. (1951). *J. clin. Endocrinol.* 11, 1021.
NELSON, D. H, SAMUELS, L. T. and REICH, H. (1951). *Proc. 2nd Clin. ACTH Conf.* 1, 49.
SAYERS, G., SAYERS, M. A., FRY, E. G., WHITE, A. and LONG, C. N. H. (1944). *Yale J Biol Med.* 16, 361.
SAYERS, G., SAYERS, M. A., LEWIS, H. L. and LONG, C. N. H. (1944). *Proc. Soc. Exp. Biol Med.* 55, 238.
SAYERS, G., SAYERS, M. A., LIANG, T-Y. and LONG, C. N. H. (1946). *Endocrinol.* 38, 1.
SELYE, H. and SCHENKER, V. (1938). *Proc. Soc. Exp Biol. Med.* 38, 518.
SELYE, H. (1949a). *Brit med. J.*, ii, 1129.
SELYE, H. (1949b). *Practitioner*, 163, 393.
STEWART, C P., ROBSON, J. S. and TOMPSETT, S. L. (1952). II^e^ *Congrès Internat de Biochem., Resumés des Communications*, p. 138.
SWEAT, M. L. (1951). *J. Amer Chem Soc.* 73, 4056.
TAIT, J. F., SIMPSON, S A. and GRUNDY, H. M. (1952) *Lancet*, i, 122.
TOMPSETT, S. L Personal communication.
VOGT, M. (1947). *J. Endocrinol* 5, lvii.
VOGT, M. (1951). *J Physiol.* 113, 129.

XI

Biochemical Genetics

A. NEUBERGER

BIOCHEMICAL GENETICS is not an independent branch of natural science; it is rather a borderland with ill-defined boundaries where the geneticist and biochemist meet. However, the frequent use of the term emphasizes the importance of biochemical techniques to the modern geneticist and the value of a genetical approach to the biochemist. It is hoped to illustrate both these aspects in the present lecture.

In the classical phase of the development of genetics a purely formal theory was developed by Mendel and no definite assumptions were either made or required about the physical or chemical nature of hereditary factors or genes. These genes, which may be defined as the irreducible units or 'quanta' of inheritance, were during the early development of genetics purely mental concepts. In the second or cytological phase of genetics it was realized that genes must be associated with certain structures in the cell nucleus and soon convincing evidence was produced that these hereditary units are contained mainly or exclusively in the chromosomes. However, even chromosomes are, chemically speaking, complex structures, containing a variety of compounds such as proteins, nucleic acid and other substances. It is a reasonable supposition that not all these substances are closely connected with the action of genes and it can be postulated that differences in gene action will be reflected in chemical differences in molecular structure of the genes. In this third and contemporary phase an attempt is being made to define the chemical nature of genes more closely and to establish by chemical and physical methods how the gene controls those

characteristics of living systems, the inheritance of which is the subject of investigation by the geneticist.

The study of gene action has produced another point of contact between the biochemist and geneticist. The latter used to examine the inheritance of such characters as eye colour or flower pigments, using purely descriptive terms. This method has often given satisfactory results, whilst in other cases the complexity of the phenomenon has defied analysis by an exclusively descriptive or morphological approach. More recently it was realized that pigment changes, for example, must be due to the occurrence of well-defined chemical reactions, and a study of these chemical changes has led to the working hypothesis that genes control, in an, as yet, unspecified manner, a particular step in a biochemical reaction chain. This approach has been particularly fruitful in microbial genetics, but it has also been *responsible for many interesting advances in our knowledge* of the genetics of higher forms of life.

The biochemist mainly interested in metabolic or even nutritional phenomena has in turn also benefited from his contact with genetics. In the first place, he has learned to realize that the potential capacity of an organism to perform certain chemical reactions is primarily determined by genetic factors. Changes in environment may bring about a great increase in the production of certain enzymes ('enzyme adaptation') but only within the framework of the genetically determined potential of the species. Thus the [illegible]
essential amino-acids [illegible]
mined by genetic factor[illegible]
the biochemist with very useful tools for studying metabolic mechanisms. Thus, as will be discussed later, the investigation of naturally occurring mutations in man has helped us towards a better understanding of the metabolism of aromatic amino-acids, and the study of artificially produced mutations in *Drosophila* and particularly in the fungus *Neurospora* has opened up new lines of research in the metabolism of a great variety of compounds.

ALCAPTONURIA

The early history of biochemical genetics is closely linked with the initial stages of development of general genetics in this country, and is particularly associated with the name of a clinician, Sir Archibald Garrod. Garrod, who at that time was on the staff of St. Bartholomew's Hospital and also on that of Great Ormond Street, became much interested, about 1898, in a rare metabolic abnormality, alcaptonuria. In this condition a urine is excreted which has strong reducing properties and turns black on standing in air. The substance responsible for this behaviour, homogentisic acid, had been isolated, analysed and characterized by Wolkow and Baumann (1891) as hydroquinone acetic acid. Garrod produced evidence that, although diagnosis may not have been made until adolescence, the condition in his and other cases was probably present from birth. He also emphasized in a paper which he gave in 1899 the fact that alcaptonuria frequently occurred in several members of a family, usually brothers and sisters, an observation already made in 1886 by another English physician, Kirk (see Garrod, 1899). It was also noted that the parents are usually apparently normal. In a second paper (1902a), Garrod drew attention to another phenomenon, the high frequency of first-cousin marriages amongst parents of alcaptonurics as compared with the frequency of first-cousin unions in the general population. At that time Mendel's discoveries, which had first been announced in 1865, were still unknown in Great Britain, although they had come to the notice of Continental biologists through the almost simultaneous publication in 1900 of three papers by de Vries, Correns and Tschermak respectively. However, William Bateson, later Professor at Cambridge, realized the importance of the rediscovery of Mendel's principles and in 1902, in a report which he made (with Miss E. R. Saunders) to the Evolution Committee of the Royal Society, pointed out that Garrod's findings can be readily explained by Mendel's work. He suggested that alcaptonuria was due to the action of a rare, abnormal gene which in the heterozygous form produced no apparent symptoms Thus each of the apparently normal parents carried one normal and one mutant gene and 25 per cent of their off-

characteristics of living systems, the inheritance of which is the subject of investigation by the geneticist.

The study of gene action has produced another point of contact between the biochemist and geneticist. The latter used to examine the inheritance of such characters as eye colour or flower pigments, using purely descriptive terms. This method has often given satisfactory results, whilst in other cases the complexity of the phenomenon has defied analysis by an exclusively descriptive or morphological approach. More recently it was realized that pigment changes, for example, must be due to the occurrence of well-defined chemical reactions, and a study of these chemical changes has led to the working hypothesis that genes control, in an, as yet, unspecified manner, a particular step in a biochemical reaction chain. This approach has been particularly fruitful in microbial genetics, but it has also been responsible for many interesting advances in our knowledge of the genetics of higher forms of life.

The biochemist mainly interested in metabolic or even nutritional phenomena has in turn also benefited from his contact with genetics. In the first place, he has learned to realize that the potential capacity of an organism to perform certain

within the framework of the genetically determined potential of

the biochemist with very useful tools for studying metabolic mechanisms. Thus, as will be discussed later, the investigation of naturally occurring mutations in man has helped us towards a better understanding of the metabolism of aromatic amino-acids, and the study of artificially produced mutations in *Drosophila* and particularly in the fungus *Neurospora* has opened up new lines of research in the metabolism of a great variety of compounds.

feeding a variety of such compounds it should be possible to establish a reaction sequence which should appear reasonable on biochemical grounds. This has been done, and as a result of many investigations it has been possible to suggest the following pathway (Fig. 1). In the first step phenylalanine is converted to tyrosine; the latter is then de-aminated to a keto acid. The next, also oxidative, step is, chemically speaking, peculiar: it consists of the probably simultaneous introduction of a second hydroxyl

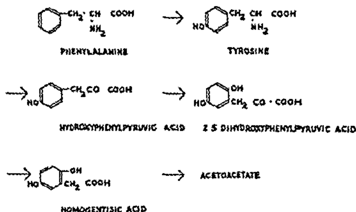


FIG 1.

group in the benzene ring and a migration of the side chain to an adjacent position in the ring. This is followed by decarboxylation to homogentisic acid. All these reactions which were originally postulated mainly on the basis of results obtained with alcaptonurics, have been amply confirmed by recent work using either isotopes or isolated enzyme systems. Thus the concept that alcaptonuria represents a genetic block has been proved. It follows from this idea that the gene involved acts in some way on the enzyme catalysing the further metabolism of homogentisic acid, by controlling either the formation of the enzyme or the activity of the enzyme in some manner, as yet unspecified. This general idea is implicit in many of Sir Archibald Garrod's writings and constitutes probably his most outstanding contribution to biological thought. However, the full significance of this concept was not realized until after his death.

spring should show the signs of alcaptonuria. As Garrod pointed out later, the figure actually found is of this order. It is also obvious that, at least with a rare recessive gene, the influence of consanguinity on the manifestation of the genetic abnormality should be very marked. Garrod, in a paper in *The Lancet* which appeared later in 1902 (1902b), fully accepted this interpretation and elaborated the wider implications of this demonstration of the effects of gene action on metabolism.

I am giving some prominence to alcaptonuria, partly because research into this condition is so closely associated with the foundation of biochemical genetics. But in addition, out of work on alcaptonuria some concepts of wider application have developed which have become of increasing importance with the passing of years. It was shown by Wolkow and Baumann (1891) that homogentisic acid arises from phenylalanine and tyrosine, and good evidence has been produced by many workers that, when either of these amino-acids is given in the diet, a complete or almost complete conversion to homogentisic acid takes place in the alcaptonuric. There are two possible explanations for the occurrence of homogentisic acid in alcaptonuric urine. It may be assumed that the breakdown of tyrosine or of the corresponding keto acid to homogentisic acid is a completely new reaction which does not occur in normal individuals; this indeed was the opinion expressed by Dakin (1911) and Grutterink and van den Bergh (1907). The alternative view, adopted by Garrod, in his Croonian lectures given in 1908 and published later (1909) as a book, is that homogentisic acid is a normal product of intermediate metabolism and one which can be further broken down by normal individuals, but not by alcaptonurics. According to this interpretation, alcaptonuria represents a genetic block in a normal biochemical reaction sequence. The concept was supported by two main lines of evidence. It is a necessary corollary of the concept of the genetic block that normal individuals should be able to metabolize homogentisic acid without difficulty and this indeed is the case. Moreover, all substances postulated as intermedi
give rise
alcaptonu

phenylalanine in the blood and in tissue fluids leads to toxic effects, particularly in the brain, and this is indeed not unlikely since an abnormally high intake in the diet of other amino-acids such as methionine or glycine has been shown to be deleterious. Alternatively, secondary products of the metabolism, such as phenylacetic acid, may be responsible for the toxic symptoms observed, especially if these substances have been present in the body already during intrauterine life. A third possibility is that the enzyme responsible for the conversion of phenylalanine to tyrosine also catalyses the hydroxylation of an unknown substance to a product necessary for the normal function of certain cortical centres.

Amino-acidurias

Cystinuria is a condition that has been known for a long time. As the name indicates, one of its important features is the excretion of relatively large amounts of cystine. The presence of considerable amounts of this very insoluble amino-acid in the urine may give rise to renal calculi in predisposed subjects. It is of interest that such a calculus was the source from which cystine—hence its name—was first isolated in 1810 by Wollaston, a distinguished chemist who, though trained as a physician, made outstanding contributions to the chemistry of metals of the platinum group and thereby acquired a large fortune. It was not until 1899 that cystine was shown to be formed by hydrolysis of proteins.

The biochemistry of cystinuria is somewhat more complicated than that of the conditions discussed so far. Garrod in his *Inborn Errors of Metabolism* already pointed out that there is strong evidence that the cystinuric can burn the sulphur of cystine to sulphate; thus, if there is a metabolic block, it is not complete. Moreover, substances other than cystine are excreted in excessive amounts in the urine. Udransky and Baumann (1889) demonstrated the presence in urine of cystinurics of two diamines, putrescine and cadaverine which can be derived from ornithine and lysine respectively. These findings have been confirmed by some workers, although many other investigators looked for these diamines in vain. But in 1912 Ackermann and Kutscher isolated lysine as its monohydrochloride from the

Phenylketonuria

Another genetically determined abnormality which is caused by a genetic block at a well-defined stage in a biochemical reaction sequence was discovered by the Norwegian scientist, Folling (1934). He found that urines of certain mentally deficient patients gave an intense colour with ferric chloride. He traced the substance responsible for this reaction and identified it as phenylpyruvic acid, the keto acid corresponding to phenylala-

with the excretion of the keto acid in the urine. It is known that almost all amino-acids are in metabolic equilibrium with the corresponding keto acids mainly through a transamination reaction with glutamic acid or α-ketoglutarate. It is reasonable to suppose that this reversible amination and de-amination between phenylalanine and phenylpyruvic acid occurs normally in phenylketonurics, but that the irreversible oxidation to tyrosine, the first step in the reaction sequence discussed above (Fig. 1), is blocked. These individuals metabolize tyrosine normally, as would be expected on this basis. The substance present in blood in excess is mainly the amino-acid which is filtered through the glomeruli, but is reabsorbed by the tubules. However, a small amount of phenylalanine gets oxidized to the keto acid and this acid is not reabsorbed, at least not to the same extent as the amino-acid, and is therefore excreted in the urine. Another, extremely rare, metabolic abnormality concerning aromatic amino-acids is tyrosinosis, which has only been observed once by Miss Grace Medes (1932). In this condition the further oxidation of tyrosine or the corresponding keto acid to 2 : 5-dihydroxyphenylpyruvic acid is apparently blocked.

One complicating feature of phenylketonuria is its invariable association with a condition of high-grade mental deficiency. The available facts suggest that both the mental defect and the metabolic 'error' are due to the action of the same recessive gene and these multiple or 'pleiotropic' effects of the gene are somewhat difficult to explain. It is possible that the high level of

are present in plasma—the quantities varying normally between about 0·3 and 4·0 mg. per 100 ml. The quantities excreted in the urine of normal man are small, and values (per day) *may* lie between 10 and 25 mg. for different free amino-acids. It is clear that only a small proportion of the amino-acids filtered by the glomeruli ultimately appears in the urine and it can be calculated that 95 to 99 per cent of the plasma amino-acids are normally reabsorbed by the tubules. No exact measurements relevant to this problem have been reported, but it would appear that in all cases of cystinuria, reabsorption of cystine still occurs. Thus with a constant plasma level of 0·8 mg. of cystine per 100 ml. and an amount of plasma filtered through the glomeruli of about 200 l./day, the total amount of cystine filtered per day is 1·6 g. The daily cystine excretion of cystinurics varies between about 0·4 and 1·2 g. It would appear therefore that in cystinuria reabsorption of cystine is only partially reduced and the same is likely to apply to lysine and arginine. The association of increased excretion of cystine and of lysine and arginine suggests that these three amino-acids are reabsorbed by the same mechanism and that this mechanism is not concerned with the reabsorption of other amino-acids. The detailed biochemical reactions involved in tubular reabsorption in general are unknown, but it is certain that such reabsorption is coupled with an energy-yielding reaction which is presumably catalysed by enzymes. It is thus probable that the action of the gene under consideration is on the enzyme or enzymes associated with the transfer of these three amino-acids across a tubular cell surface and that in cystinuria the activity of such enzymes is decreased resulting in less efficient reabsorption.

It is not possible to deal in detail with other amino-acidurias such as those associated with hepatolenticular degeneration (Wilson's disease) or the rather complex Fanconi syndrome. In these conditions, particularly in the Fanconi syndrome, the excessive amino-acid excretion in the urine is also of renal origin, but in contrast to cystinuria, the urinary levels of *all* amino-acids are abnormally high. These conditions are clinically complex and defy at present a complete biochemical interpretation.

Before we leave 'inborn errors of metabolism', I should like to

urine of a cystinuric and later Hoppe-Seyler (1933) obtained relatively large amounts of another derivative of this amino-acid. Yeh *et al.* (1947) using microbiological methods reported the presence of excessive amounts of cystine, arginine and lysine in the urine of cystinurics. More recently Dent and Rose (1951) in this country, and Stein (1951) in the United States, applying modern chromatographic techniques due to Martin and Synge, showed that the urine of cystinurics contains, apart from cystine, relatively large amounts of lysine invariably and, somewhat irregularly, of arginine; other amino-acids were found to be

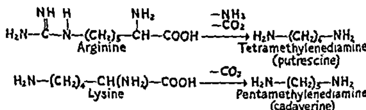


FIG. 2.

present in normal concentrations. It is thus possible that the two diamines found occasionally in urines of cystinurics arise from

a number of amino- and diamino-acids in common'.

It is somewhat unlikely, in view of the structural dissimilarity between cystine and lysine, that cystinuria should consist of the type of metabolic block encountered, for example, in alcaptonuria. It occurred, independently, to various workers a few years ago (see particularly Dent, 1949) that this condition is more readily explained if it is assumed to be essentially renal in nature. This was proved conclusively by Fowler, Harris and Warren (1952) who showed that the concentration of cystine in the blood of cystinurics was normal. The high concentration in the urine cannot therefore be due to a metabolic disturbance in the usual sense, but must be caused by some derangement of kidney function.

Almost all the amino-acids obtained by hydrolysis of proteins

is oxidized by an enzyme-complex to formylkynurenine, a reaction first demonstrated by Knox and Mehler (1950). In the second step the formyl group is removed by another specific enzyme present in mammals in the liver. The resulting kynurenine is then hydroxylated to give hydroxykynurenine which is split to 3-hydroxyanthranilic acid by an enzyme containing pyridoxal phosphate as a coenzyme. The mechanism by which the latter is converted to nicotinic acid is still uncertain.

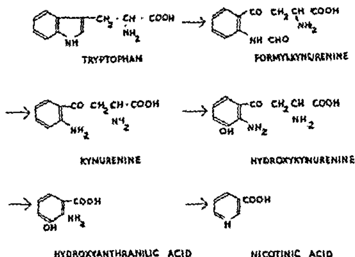


Fig. 3.

It is impossible within the scope of the present lecture to produce detailed evidence for this chain of reactions, and I shall have to restrict myself to the results obtained with *Neurospora*. Beadle and Tatum (for the earlier work, see review by Beadle, 1945) had been producing since 1942, by X-ray treatment or less frequently by irradiation with ultra-violet light, a series of 'nutritional' mutants of *Neurospora* which have become particularly interesting to the biochemist. The wild type of *Neurospora* is very modest in its growth requirements and does not need a supply of either nicotinic acid or tryptophan in its medium. However, after X-ray treatment several strains were obtained which did not grow on a simple medium, but required nicotinic

say something about the formal genetics of these conditions. Phenylketonuria is definitely homozygous, i.e. it is due to the operation of a recessive gene, and so are most cases of alcaptonuria. There are, however, a few reports in the literature, albeit not well documented with respect to biochemical findings, suggesting that another genetic type of alcaptonuria exists which is due to the operation of a dominant gene. Cystinuria has generally been considered to be caused by a recessive gene, but according to a recent paper by Dent and Harris (1951) this seems less certain. It is possible that cystinuria is a heterozygous condition with incomplete manifestation or that genetically two types of cystinuria exist. The genetical, clinical and biochemical analysis of the Fanconi syndrome and allied conditions is more complex and this syndrome is most likely to represent a heterogeneous mixture of various deficiencies of renal reabsorption.

Tryptophan Metabolism

A field of biochemistry in which the application of genetic techniques together with more conventional biochemical methods has produced striking advances is the metabolism of tryptophan. It has been known since its discovery by Hopkins and Cole more than fifty years ago that tryptophan is an essential constituent of the diet for mammals and it was generally assumed that it was mainly or exclusively required as a building stone of proteins. In 1945 Elvehjem and his colleagues in Wisconsin showed that tryptophan had a second and an important nutritional function which consisted in being a precursor of nicotinic acid. The rather complicated chemical reactions which bring about this conversion have been elucidated by the use of a great variety of methods, but the application of mutant strains of the bread mould *Neurospora crassa* has played a most valuable part. These investigations are important in two respects. In the first place they have demonstrated in interplay between two groups of essential dietary constituents, viz. amino-acids and at least two vitamins, pyridoxin and nicotinic acid. Secondly, they have shown that the metabolic pathway of tryptophan is on the whole similar in mammals, insects, moulds and bacteria. The main series of reactions is shown in Fig. 3. In the first step tryptophan

substitution of a normal gene by its mutant allele is associated with the apparent failure of a definite chemical reaction to take place and frequently with the accumulation of a substance which occurs immediately before the 'blocked' step in a sequence of reactions. There is the further important point that, at least with nutritional mutants, substances which are put in the postulated reaction scheme before the 'blocked' step, are not utilized, whilst substances which are placed after the 'genetic block' are active. These findings led Beadle in 1945 to put forward the hypothesis that there is a unique correspondence between genes and enzymes. He assumed that each gene acted specifically on one enzyme and that a loss of a gene necessarily leads to the loss of the corresponding enzyme. This stimulating and attractive hypothesis has been interpreted in various ways and it has even been suggested that genes themselves are enzymes; but this seems most improbable. More frequently it has been assumed that enzymes are direct gene products, leaving the question open as to how enzyme synthesis is governed by the specific gene. Anyhow, this hypothesis in its original form can no longer be maintained, although it has been useful in interpreting so many observations in biochemical genetics.

It has been known for some time that mutant strains of *Neurospora* could be produced which responded either to tryptophan or nicotinic acid. If a genetic block is complete and if the reaction sequence discussed above is correct, such strains should not exist. In order to explain these surprising findings, Bonner and his colleagues (see Bonner, 1951) have carried out extensive investigations using ^{15}N as tracer. Their results have clearly shown that reactions which are apparently blocked, as shown by lack of growth, can still take place in the mutant strain to a limited extent. Thus, mutants which are apparently blocked in the synthesis of tryptophan, can grow if sufficient nicotinic acid is supplied. These mutants cannot synthesize enough tryptophan to cover both the requirement for the amino-acid itself and for nicotinic acid as well. The observation that certain strains respond to either tryptophan or nicotinic acid is thus explained. This phenomenon of a quantitative rather than a qualitative effect of genes on enzyme and which has been termed 'leakage' indicates

acid or one of its precursors. The results of a systematic analysis of the nutritional requirements of these mutant strains were compatible with the hypothesis that each of these mutations consisted of the loss of the capacity to carry out one particular step in the reaction chain (see review of Mitchell, 1950). Thus one strain for which the nicotinic acid requirement could be replaced by tryptophan grew well on kynurenine, hydroxykynurenine and hydroxyanthranilic acid. Another strain labelled N-3 was found to be unable to use kynurenine but could use hydroxyanthranilic acid. Culture filtrates from this strain were found to support the growth of another mutant, N-4. The active substance accumulated by N-3 was isolated and identified as *N*-acetyl kynurenine. N-3 obviously cannot hydroxylate kynurenine and accumulates the latter as acetyl derivative. Similarly another strain, N-2, accumulated hydroxyanthranilic acid and is presumably unable to carry out the further reactions leading to nicotinic acid.

Another example of gene action of tryptophan metabolism is provided by the work of Kuhn and Butenandt, which has been recently reviewed (Butenandt, 1952). In the wild race of *Drosophila melanogaster* and of *Ephestia kuhniella* (flour-moth) the eyes are dark in colour and this colour is produced by the accumulation of certain substances called omnochromes. Three different genes at least are concerned in the formation of these pigments. In one mutant strain, normal eye colour is obtained, if kynurenine is supplied and it seems certain that the ability is lost to convert tryptophan to kynurenine. But there is a a second gene which intervenes at a later stage and controls the conversion of kynurenine to hydroxykynurenine. The latter is further changed to another unknown product which is linked to a protein carrier and gives rise to the actual pigment. At least one gene is associated with this last stage. In this example genetic analysis together with the use of biochemical methods has, at least partly, elucidated the mechanism of pigment formation.

The Interrelationship between Genes and Enzymes

In alcaptonuria and phenylketonuria, in the examples given of mutations in *Neurospora* and insects and many other cases,

of the mutant protein differs from that of the normal, leading to a change of solubility which in turn is responsible for the morphological changes. This important finding indicates that genes may have a direct effect on protein synthesis.

In conclusion it appears that there is close connection between genes and the production and activity of certain proteins, particularly enzymes. The simple hypothesis that each gene controls the production of one enzyme and conversely that the formation of each specific protein or enzyme is catalysed by a special gene is probably too simple. Many genes may act in this way, but the available evidence suggests that in other cases gene action may be more indirect. Only a combined attack using both chemical and genetic methods is likely to produce further progress in this interesting and important field.

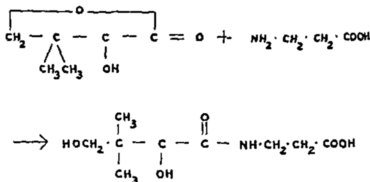
The Chemical Nature of the Gene

As already mentioned, convincing cytological evidence was obtained many years ago indicating that only chromosomes are concerned with the transmission of Mendelian characters. This soon suggested that the deoxyribonucleic acids (DNA) or deoxyribonucleic acid-protein complexes, both of which are found only in the nucleus and not in the cytoplasm, might be identical with genes or form an important part of a more elaborate gene structure. Until recently, the relatively few experimental observations bearing on this point were somewhat overshadowed by an extensive speculative literature. However, in the last decade two important discoveries were made which greatly strengthen the hypothesis that genes are structures containing DNA and that the latter largely determines the specificity of the gene.

Griffith (1928) had shown that non-encapsulated pneumococci could be made to synthesize a capsule of certain type under the influence of some substance in the capsulated pneumococci. The change produced was permanent, i.e it was transmitted to the descendants of the transformed organisms. In 1944, Avery, Macleod and McCarthy reported that this material was a deoxyribonucleic acid and in 1945 Boivin *et al* reported similar findings for certain strains of *E. coli*. The preparations of

that genes, whilst influencing or controlling enzymic reactions, do not necessarily determine presence or absence of an enzyme.

Another observation is also relevant to this problem. Wagner (1949) has described a mutant strain of *Neurospora* which is incapable of synthesizing pantothenic acid from β-alanine and α-hydroxy β-dimethyl γ-butyrolactone (pantoyl lactone) (Fig. 4). This reaction can be performed by the wild strain. It was found, however, that extracts of both types of strain were capable of



PANTOTHENIC ACID

FIG. 4

promoting the synthesis of the vitamin from the two fragments. In this instance the mutant gene definitely did not inhibit the formation of the enzyme in question, but prevented it from being active *in vivo*. The question as to how this gene affects the action of this enzyme is still open.

One of the most striking effects of gene action on molecular

[illegible]

amongst negroes. The anaemia which is quite severe occurs in the homozygous condition whilst in the heterozygous state the sickling trait is found. It was observed that the haemoglobin in the sickle-cell anaemia patients differs in isoelectric point and solubility but not in amino-acid composition from normal haemoglobin. It appears that the configuration or conformation

The opinion is widely held that ribonucleic acid, which is found mainly in the cytoplasm, is specifically concerned with protein synthesis. If we accept the hypothesis, discussed in the earlier part of this lecture, that genes control the formation of specific enzymes and if we also subscribe to the concept that genes are deoxyribonucleic acids or DNA-complexes, it would follow that DNA too is involved in protein synthesis. No direct or conclusive evidence of the role of either type of nucleic acid in the formation of proteins is yet available. Further progress in this field is likely to arise from the combination of chemical and biological techniques; in the meantime we may give way, within reason, to the temptation to speculate on these important topics, largely in the hope that such speculations may stimulate the design of further, conclusive, experiments.

REFERENCES

ACKERMANN, D. and KUTSCHER, F. (1912). *Z. Biol.* **37**, 355.
AVERY, O., MACLEOD, C. M. and MCCARTHY, M. (1944). *J. Exp. Med.* **79**, 137.
BATESON, W. and SAUNDERS, E. F. (1902). Report to the Evolution Committee of the Royal Society No. 1, p 33, footnote
BEADLE, G. W. (1945). *Chem. Rev.* **37**, 15.
BOIVIN, A., DELAUNAY, A., VENDRELY, R. and LEHOULT, Y. (1945). *Experientia*, **1**, 334.
BOIVIN, A., VENDRELY, R. and VENDRELY, C. (1948) *C. R. Acad. Sci.* (Paris), **226**, 1061
BONNER, D. M. (1951) *Cold Spr Harb Symp quant Biol* **16**, 143
BUTENANDT, A. (1952) *Endeavour*, **11**, 188
DAKIN, H. D. (1911) *J biol. Chem* **9**, 151
DAVIDSON, J. N. and LESLIE, I. (1950). *Nature*, **165**, 49
DENT, C. E. (1949) *Biochem Soc. Symposia*, No. 3
DENT, C. E. and HARRIS, H. (1951) *Ann. Eugen.* **16**, 60.
DENT, C. E. and ROSE, G. A. (1951). *Quart. J. Med.* **20**, 205
FOLLING, A. (1934). *Hoppe-Seyl Z* **227**, 169.
FOWLER, D., HARRIS, H. and WARREN, F. L. (1952). *Lancet*, **1**, 544
GARROD, A. E. (1899). *Medico-Chirurg Trans.* **82**, 367
GARROD, A. E. (1902a). *Medico-Chirurg Trans.* **85**, 69.
GARROD, A. E. (1902b) *Lancet*, **ii**, 1616.

DNA used contained only minute amounts of protein and the activity of the material was destroyed by the enxyme deoxyribonuclease. There can therefore be little doubt that the action of the extracts was due to the nucleic acid itself. More recently Hotchkiss (1951) reported that resistance to penicillin in pneumococci could also be transferred to non-sensitive strains by the appropriate DNA-fractions. It is thus clear that a mutation can be produced by a specific nucleic acid.

A unique correlation between quantity of chromosomal material and amount of DNA per cell has recently been demonstrated with the aid of a variety of analytical techniques. It was first shown by Boivin, Vendrely and Vendrely (1948) that for a particular species the amount of DNA per diploid cell was constant, independent of the source of the tissue and twice that found for the haploid sperm or egg. These findings have been confirmed and extended by many workers (see e.g. Mirsky and Ris, 1949; Davidson and Leslie, 1950; Vendrely and Vendrely, 1949; Schrader and Leuchtenberger, 1950). Differences of DNA content per cell between different species and phyla are sometimes quite large, but the values for the same species are relatively constant. Apparent deviations are largely explained by the fact that many cells such as those of the mammalian liver are polynucleate or that many nuclei are polyploid, i.e. the number of chromosomes per nucleus exceeds the diploid value. In other cases such as the salivary glands of *Drosophila* each chromosome is composed of a large number of threads or chromonemata ('polyteny'); but even so the correlation between DNA content and number of chromosome threads appears to be maintained (Kurnick and Herskowitz, 1952). This quantitative correlation greatly strengthens the hypothesis that DNA is uniquely associated with gene action. It would be premature and possibly even wrong to assume that the specificity of genes is in all cases entirely due to deoxyribonucleic acid. But the facts so far established suggest a unique role of DNA in gene action and specificity; they also require a great complexity in the structure of DNA and particularly a variability in its composition or intramolecular linkages. This concept is in accord with our more recent knowledge of the chemistry of nucleic acids.

XII

Chromatography in the Study of Amino-Acid Metabolism

C. E. DENT

In this lecture I propose to deal first with amino-acid metabolism briefly and in a general manner. I shall limit this to information that has been obtained by an older but still valuable method of analysis, namely for total α-amino-nitrogen. This method gives one figure only representing all the amino-acids added together. It tells us nothing of the level of individual amino-acids in the fluid analysed I shall then describe very briefly three of the new chromatographic methods that have been introduced in the last few years and which enable us for the first time to detect, isolate and determine quantitatively the individual amino-acids present in crude biological fluids Finally, I shall describe how these new methods have been applied to one small problem of practical and theoretical interest, namely, to the excretion of amino-acids in the urine.

Amino-Acid Metabolism

Fig. 1 summarizes the information obtained by amino-nitrogen assays concerning the fate of amino-acids in the mammalian organism. It is to be noted that the main source of amino-nitrogen is in the form of dietary protein. In the gut this protein is broken down into its constituent 20 or so amino-acids. The enzymes which do this act with most remarkable speed and efficiency, breaking down the protein piece by piece. The gut wall appears to be impermeable to all but traces of the more complex peptide fragments, so that most if not all of the

GARROD, A. E. (1909). *Inborn Errors of Metabolism.* Oxford University Press, London.
GRIFFITH, F. (1928). *J. Hyg.* **27**, 113.
GRUTTERINK, A. and HIJMANS V. DEN BERGH, A. A. (1907). *Nederl. Tijdschr. v. Geneesk.* ii, 1117.
HOPPE-SEYLER, F. H. (1933). *Hoppe-Seyl. Z.* **214**, 267.
HOTCHKISS, R. D. (1951). *Cold Spr. Harb. Symp. quant. Biol.* **16**, 457.
JERVIS, G. A. (1947). *J. biol. Chem.* **169**, 651.
KNOX, W. E. and MEHLER, A. H. (1950). *J. biol. Chem.* **187**, 419
KURNICK, N. B. and HERSKOWITZ, I. H. (1952). *J. Cell. comp. Physiol* **39**, 281.
[illegible]
[illegible]
[illegible] 27.
PAULING, L., HANO, H., SINGER, S. J. and WELLS, I. (1949). *Science,* **110**, 543
SCHRADER, F. and LEUCHTENBERGER, C (1950). *Proc Nat. Acad. Sci.* **35**, 464.
STEIN, W. H (1951). *Proc. Soc. exp. Biol. Med.* **78**, 705.
UDRANSKY, VON L. and BAUMANN, E. (1889). *Hoppe-Seyl. Z* **13**, 562.
VENDRELY, R. and VENDRELY, C (1949) *Experientia,* **5**, 327.
WAGNER, R. P. (1949). *Proc. Nat. Acad. Sci. Wash.* **35**, 185.
WOLKOW, M and BAUMANN, E. (1891) *Hoppe-Seyl. Z* **15**, 228.
YEH, H. L., FRANKL, W., DUNN, M. S., PARKER, P, HUGHES, B and GYÖRGY, P. (1947). *Amer. J. Med. Sci.* **214**, 507.

much faster than it can be absorbed. It is because this digestive mechanism is so efficient even in severe disease states that there is rarely any advantage in feeding proteins in pre-digested (hydrolysate) forms. Perhaps the most interesting mechanism is that by which the cells take up amino-nitrogen from the surrounding tissue fluids. This is an active process requiring energy, for the amino-nitrogen is concentrated 5–10 times in the cell presumably to facilitate protein synthesis and other amino-acid reactions. Two further processes are also shown in Fig. 1. The first is a small leak of amino-nitrogen from the blood into the urine. It is so small, only comprising 1–2 per cent of the daily intake of amino-acids, that I have shown the arrow as a dotted line. The second process is an important one that involves only the pregnant female. It concerns the transplacental passage of amino-nitrogen from the maternal blood into the foetus. This is quantitatively interesting since the placenta appears to be able to 'pump' amino-nitrogen from the mother into the foetus so that foetal systemic blood has always a higher concentration than maternal. In the human the foetal-maternal ratio is about 1·5–2·0, in guinea-pigs it is fivefold.

This brief scheme already can be related to certain matters of clinical importance. In severe pancreatic fibrosis the proteolytic enzymes are lessened in amount and inefficient protein absorption may occur. A similar state of protein deficiency can occur in severe diarrhoea, the malabsorption being due to the protein and its hydrolytic products passing too rapidly through the small intestine. In liver disease the liver cells cannot take up and metabolize the amino-acids normally, in cancerous growth they appear to do it too well. In toxaemia of pregnancy there is some evidence that the placental amino-acid pump is impaired. Finally there are certain rare diseases (I shall say more of this soon) characterized by an excessive loss of amino-acids into the urine owing to the presence of a malfunctioning kidney. I must stress however the grave errors of assuming that the amino-nitrogen figure allows us to speculate far along these lines. It does not. Arguments based on amino-nitrogen determinations all assume that the amino-acid mixture comprises all the common amino-acids in fairly constant proportions, so that a

absorption into the portal blood is in the form of free amino-acids. From here the amino-nitrogen reaches the liver where some is taken up by the cells while some passes through into the systemic blood to reach the other tissues. When absorbed into the liver and tissue cells the amino-acids enter into the various

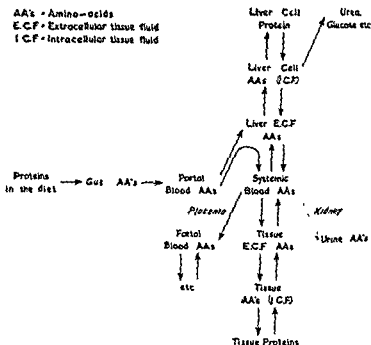


FIG. 1 Metabolism of amino-nitrogen in the mammal. This information has been obtained by assays for total α-amino-nitrogen.

reversible reactions concerning the formation of protein constituents. In some sites however they are irreversibly metabolized, the nitrogen forming urea and the carbon and hydrogen atoms being burnt eventually to carbon dioxide and water. Our amino-nitrogen analyses tell us more about the interesting details. During a heavy protein meal the blood level, normally about 4 mg per cent, rises only to about 6–8 mg per cent so efficient are the tissue cells, especially those of the liver, at removing extra amino-nitrogen from the circulating plasma. The gut contains far higher concentrations, digesting the protein

so even if these small quantities of each amino-acid are present together with relatively large quantities of many other commonly occurring substances (salts, protein, urea, glucose, etc.); it must at the same time give a roughly quantitative estimate of the amount present; it must detect hitherto unknown amino-acids (if present) as well as known ones; finally it must obtain

paper chromatography devised in 1944 by Martin and his co-workers (Consden, Gordon and Martin, 1944). In this method large sheets of filter paper are used for the analysis. The apparatus comprises a large airtight cabinet containing across the upper part a long trough which is filled with solvent. One edge of the sheet of filter paper dips into the trough, the rest of it hanging down. This enables the solvent to soak slowly down the paper. The process is shown diagramatically in Fig. 2. The square represents the sheet of filter paper. Before insertion into the cabinet the amino-acid containing fluid is applied to the paper at a point near one corner. In the example shown the right edge is then dipped into the trough which contains phenol. When it has soaked almost to the far edge of the paper it is dried and reinserted with its bottom edge into collidine solution. It is again dried after the second solvent has soaked across the paper. It is finally sprayed with ninhydrin solution and warmed. The final chromatographic analysis is then revealed to us It comprises the sheet of filter paper which is now covered with purple spots. Each spot represents microgram quantities of an amino-acid, its size and colour strength give a rough measure of the amount present. Its final position on the paper indicates its identification, as under standard conditions each amino-acid always migrates to the same position. I must mention in passing that a simpler method using only one solvent and a strip of filter paper can also be used However the comparative lack of specificity should be obvious from the diagram (Fig. 2). The four amino-acids chosen separate well from each other on the square chromatogram. When however the same mixture is run with the same two solvents on strips of paper only, overlapping will occur

doubling of the figure means a doubling of each individual amino-acid. It is surprising in retrospect that correct conclusions have so often been arrived at using such dangerous and unlikely assumptions. We know now that the amino-acids must be considered as individual metabolites and not classed together as worthy of only collective mention. In particular 10 of the 20 common ones have been singled out by W. C. Rose as 'essential amino-acids', substances which are necessary in the diet and are irreplaceable by any other substances. Let us turn back to my scheme of amino-nitrogen metabolism (Fig. 1). Surely protein synthesis in the cell will only be facilitated by an active cellular concentrating mechanism for amino-acids if all the amino-acids required for protein synthesis are equally concentrated? Similarly the placental amino-nitrogen pump will only facilitate the growth of the foetus if it pumps all the amino-acids equally. Finally the small loss of 1–2 per cent of amino-nitrogen in the urine would be quite important to the organism if it were all derived from a single 'essential amino-acid' such as methionine.

I hope it is now clear that for a further elucidation of the problems of amino-acid metabolism new analytical methods are required. It is hardly surprising that we have been slow in developing this field in comparison with, for instance, that of carbohydrate and fat metabolism. The difficulties are immense. Amino-acids never comprise more than a very small proportion of the total solids of a biological fluid. Moreover all 20 or so of them are always present together and since most of them have almost identical chemical and physical properties, the analysis of individual constituents becomes almost beyond hope of achievement. Nevertheless these problems have been solved recently by the use of chromatographic methods and I will now proceed to a brief account of the commoner and more useful techniques.

Chromatographic Techniques

The first requirement of a more detailed approach to the study of amino-acid metabolism is a good specific method of qualitative analysis The method must tell us which individual amino-acids are present in any biological fluid; it must do so using only minute microgram quantities of materials, it must do

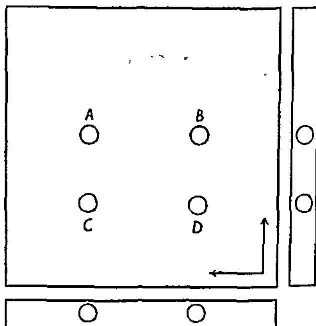
chromatograms, the one on the left shows all the amino-acids as revealed in the usual way, that on the right shows the same mixture analysed in the same way except that a minute trace of copper carbonate has been rubbed along the paper. This forms complexes with the α-amino-acids but not with 'non-α' ones. The latter therefore are the only ones to survive this treatment. When one is in doubt therefore as to the identification of an amino-acid which runs to a position close to both an 'α' and a 'non-α' amino-acid, the 'copper trick' readily enables one to differentiate between them (Crumpler and Dent, 1949). There are very many more methods of this kind.

Early work using paper chromatography showed that many amino-acids occurred in biological fluids whose behaviour could not be matched by the known common amino-acids. A method was therefore required to isolate and purify these compounds for final identification by classical chemical methods. This difficult task was largely solved in 1949 by Partridge and Westall using a displacement chromatographic method with ion-exchange

[illegible]
say that it plays an essential role in studies of amino-acid metabolism, for speculation is worthless without in the first place a full knowledge of the actual metabolites that may be present.

The final technical requirement is that of an accurate quantitative micro-analysis of all the amino-acids present in a biological fluid. This has recently been provided for us by the herculean labours of Moore and Stein (1951). They use an ion-exchange resin in the manner of an elution chromatogram. The fluid is poured on top of the column which is then eluted with various buffered fluids at different pH's. The amino-acids are washed down the column in the form of bands, at different rates, and therefore emerge separately one after the other from the lower end of the column. An automatic drop counting device collects small fractions of constant volume in the series of test tubes. The tubes are eventually analysed colorimetrically for *amino-nitrogen and a diagram drawn relating qauntity of* amino-nitrogen in each few ml. of eluate. The diagram takes the

and the mixture will appear to contain only two substances. The amino-acid identifications are of course made by first determining the movements of known pure compounds and plotting them on a reference map. About a hundred amino-acids have been plotted to date in this way in our laboratory. Most of them



if present in biological fluids can be identified simply by position matching against the reference map In certain cases there is ambiguity on account of the closeness of the positions. Then additional methods must be used. I show one of these, the 'copper trick', in Plate XV, Fig. 3.[1] Here are two finished

[1] The plates referred to in the course of this lecture will be found at the end of the book.

We can now fulfil our requirements for a new approach to the study of amino-acid metabolism. We can detect the individual substances present, identify new compounds as and when found, and finally we can determine the amounts of each quantitatively. Let us now give a few examples of the applications of these methods. These examples have to be limited somewhat; they will stress as much as possible general and medical problems, and they will have to deal mainly with the results of paper chromatography since this is the only method that has been available for a sufficient length of time.

Chromatograms of Body Fluids

I shall now briefly consider some representative chromatograms of body fluids. The first is an analysis of urine from a normal human (Plate XVI, Fig. 5). With the volume taken here only a few of the large number of amino-acids present are revealed as spots. The strongest is that due to glycine, the weaker spots comprise alanine, glutamine, taurine, etc. The first thing to note is that the chromatogram tells us not only what amino-acids are present, but also their approximate relative proportions. It reveals a new world of scientific investigation. Furthermore the picture can be memorized and is, to this extent, better than a row of figures indicating exact quantities of amino-acid. Plate XVI, Fig. 6 shows the picture obtained from human plasma. It is to be noted that the picture shows an entirely different pattern of spots from the urine chromatogram (Fig. 5). As these specimens of plasma and urine were both taken from the same subject at approximately the same time, it is clear that the kidney must be the organ responsible for the change in pattern. The next plate (XVII, Fig. 7) shows an analysis of cerebrospinal fluid, the volume taken being the same as that of the plasma previously. The pattern is again quite different, only one spot, that due to glutamine, being of comparable strength to that from plasma, all the other amino-acids being present in much lower concentrations. Clearly cerebrospinal fluid cannot be a simple ultrafiltrate of plasma. Plate XVII, Fig. 8 shows an analysis of human sweat. This is a most interesting fluid rich in amino-acids, and with several differences from plasma, a unique

form shown in Fig. 4. Each wave represents each amino-acid. It is identified by the volume of eluate required to wash it out of the column. It is determined quantitatively by the total of amino-nitrogen present in the fraction. This method can determine all the amino-acids in a mixture on a microgram scale. It

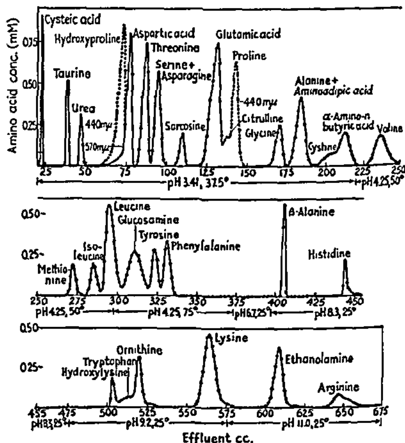


Fig. 4 Elution diagram of one of Moore and Stein's amino-acid assays.

takes up to two weeks' continuous running to complete an analysis, but this is a small price to pay for such an accurate specific method. The recoveries obtained by Moore and Stein on analysis of synthetic mixtures have shown that a high order of accuracy is obtainable.

urine analyses have shown that other amino-acid patterns are also found in normal people, always being constant in the given individual, and always being formed from plasma of an apparently normal composition. It began to look as if one's urine amino-acid pattern were part of one's individuality like one's fingerprint or other body constants. Such speculations have

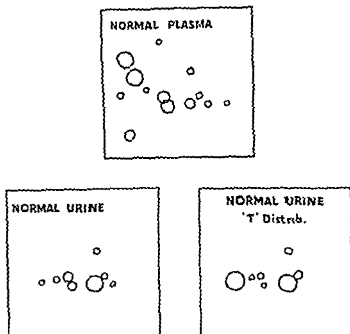


FIG. 9. Diagrams of chromatograms from normal plasma (625 μl) and from two normal urines. The size of the circles gives some indication of the strengths of the corresponding spots.

led to further work by Datta and Harris (1951) on animal urine. Their analyses abundantly proved the thesis. Different animals also excrete different urine amino-acid patterns, each pattern being constant for the animal. The few plasmas analysed however were surprisingly similar to human plasma. This may be illustrated by three chromatograms from animal urines. Plate XVIII, Fig. 10 shows the urine of a mouse. This is characterized by a very large excretion of taurine, the other amino-

and spectacular one being the strong spot due to citrulline. Sweat is the only fluid so far investigated which contains such high relative proportions of citrulline. We have no idea yet of its function there.

I think it must be already apparent that the amino-acid patterns of various body fluids may be quite different from each other. This must largely destroy my earlier scheme (Fig. 1) of amino-acid metabolism which was based only on determinations of α-amino-nitrogen. The chromatographic studies tell us a great deal more. Plasma, cerebrospinal fluid and sweat amino-acid patterns are very different from each other, yet plasmas from different people, and, as far as has been investigated, from other mammals, resemble each other fairly closely (as if they constituted a true 'milieu intérieur'), and the few facts so far accumulated

except urine.

in particular that I shall devote the remainder of my lecture.

Amino-Acids in Urine

Fig. 9 shows diagrammatically the amino-acid pattern for plasma together with the commonest urine pattern on the bottom left. These patterns are the same as those photographed in Plate XVI, Figs. 6 and 5 respectively On the bottom right of Fig. 9 however is shown another pattern characterized by the excretion of relatively large quantities of an amino-acid present only in traces in the first urine. This urine pattern has been constantly excreted in the urine of one normal individual for at least four years. His plasma is normal however. He must therefore have a different kidney function from the other normal human in that the clearance for this amino-acid is relatively higher. We were particularly intrigued by the fact that we could not identify this substance except to show by the 'copper trick' that it was a 'non-α' amino-acid. It had therefore to be isolated by the ion-exchange resin method and then analysed and characterized. It was shown to be β-amino-isobutyric acid, an amino-acid not previously described from natural sources and of particular interest to the biochemist because of its possessing a short branched carbon chain (Crumpler, Dent, Harris and Westall, 1951). Further

a new and important factor is introduced, namely, some of these patterns are associated with more or less severe disease.

Plate XIX, Fig. 13 shows the urine amino-acid pattern of a patient with the Fanconi syndrome. There is a large excretion of very many amino-acids again occurring in the presence of normal blood levels of total α-amino-nitrogen and of individual amino-acids as determined chromatographically. Here however the abnormally high renal clearance of amino-acids, which this implies, is associated with other signs of kidney tubule damage. This is shown by the presence also of renal glycosuria, inability to pass an acid urine or make much ammonia, and often also defective tubular reabsorption of phosphate, water and potassium. The patient gets rickets or osteomalacia as well as symptoms due to the other electrolyte defects. The urine amino-acid pattern is as before constant under a wide range of dietary and environmental conditions. We have some evidence also (Dent and Harris, 1951) to suggest that the pattern is lifelong and may therefore precede the onset of the clinical syndrome as if it were very closely linked with the action of the abnormal gene which the patient has inherited.

Plate XX, Fig. 14 shows the chromatogram from the urine of a patient who has been troubled by renal stones which on analysis were shown to be made of cystine (Dent and Rose, 1951). The large spot indicates the excessive quantity of cystine present in the urine. There are however two more strong spots, one due to arginine and the other to lysine and ornithine (which overlap on the chromatogram). The lysine excretion is actually greater than that of the cystine. Again we have a normal blood concentration of amino-acids but, as is not the case in the Fanconi syndrome, there is no other kidney defect present unless the patient has had complications from stone formation. Hence the cystinuric is not to be considered as a diseased individual. He is better considered as a normal person who has a tendency to stone formation If he is lucky and gets only few or no stones he will live as long and as normally as anybody else. The literature is confused on this subject and the stone-forming cystinurics have been sometimes considered as suffering from the same disease as those who have the Fanconi syndrome. The clinical

acids being relatively so much weaker as to be hardly visible. Plate XVIII, Fig. 11 shows the analysis of the urine of the domestic cat. It shows a strong spot due to methyl histidine and traces of other substances, the most interesting being a new sulphur-containing amino-acid which Westall is in the course of identifying and which we have called 'cat-spot' provisionally. Plate XIX, Fig. 12 is from the urine of the blotched genet (*Genetta tigrina*), a wild cat from Kenya. The main spot here is due to cystine. This cat has a gross cystinuria, the urinary cystine concentration being about twice that of a human cystinuric, yet as far as we know the genet never gets cystine stones. We are in course of investigating this interesting problem.

Before going farther in the observations of urinary amino-acid pattern a few notes of caution are in place. It must first be stressed that a pattern that is constant according to only a very rough quantitative method, such as paper chromatography, may in fact be found to be rather inconstant when more accurate methods of analysis are applied. Moreover the above claims have so far only been made in the case of those relatively few amino-acids which are the main constituents of urine. No attempt has been made yet to study the pattern when more urine is analysed on the chromatogram so that more amino-acids become visible. Finally in normal people the pattern does clearly vary with dietary intake in the case of those few amino-acids which have a high renal clearance. The best examples of this are histidine and methylhistidine which may greatly increase in output after a heavy protein meal, presumably because their high clearances ensure that minor changes in blood level cause proportional changes in urine output. Furthermore in patients with gross renal amino-aciduria (see below) it is particularly easy to alter the urine pattern by protein feeding since most of the amino-acids are being cleared rapidly from the blood. In such cases the importance of taking fasting urine specimens must therefore be borne in mind when the pattern is being studied for diagnostic purposes.

I shall now consider some more chromatograms from human urine. Up to now the urine amino-acid patterns have been considered to be interesting normal variants. From now on however

of normal blood amino-acid levels. I have no doubt that many more new syndromes will be discovered which are also characterized by an abnormal amino-acid excretion.

To summarize this data. We have seen that various individuals both normal and diseased have constant differences in

TABLE 1. Classification of 'Cystinuria' when this Term is used to indicate not a particular Syndrome but rather any condition characterized by an increased excretion of Cystine into the Urine

	Amino-acids found in urine in abnormally large amounts	Clinical syndromes
1. RENAL CYSTINURIA (Normal plasma cystine level)	(a) Cystine + lysine ± arginine.	Normal health usually. Occasional cystine stones and renal damage.
	(b) All the common amino-acids and cystine + lysine ± arginine	Lignac's disease (Cystinosis). Debré-deToni-Fanconi syndrome.
	(c) As (b) above.	Hepato-lenticular degeneration (Wilson's disease).
2. HEPATIC CYSTINURIA (Raised plasma cystine level)	(a) Cystine alone	Mild liver damage (of any cause).
	(b) Cystine and 2–3 other rare amino-acids (*no* lysine)	More severe liver damage.
	(c) Cystine and all the other common amino-acids	Acute yellow atrophy.

kidney function which lead to variations in the urine amino-acid patterns which occur in the presence of normal blood levels. To complete the picture I must add that there are also chronic and temporary conditions that I shall not deal with in this lecture in which the urine pattern can change in association with and presumably as the result of changes in plasma patterns. These latter conditions (but not the former) are correctly called errors of amino-acid metabolism since the blood homeostatic mechanisms are disturbed. A chronic congenital condition of this kind is

confusion is easy to understand in the case of a cystinuric who happens to have renal damage and maybe renal rickets too. Moreover cystine excretion is increased in both diseases. The chromatograms however readily distinguish between them, the urinary patterns being quite different and in the case of stone-forming cystinurics they are so characteristic that the diagnosis can be reliably made by post from analysis of a drop of urine and without ever needing to see the patient. The paper chromatographic analyses have now been supported by accurate analyses using the ion exchange resin method. Stein (1951) has analysed the urine from five cases of cystinuria and has shown quantitatively how closely the relative quantities of cystine, arginine, lysine and ornithine resemble one another. There appear however to be two sub-groups with two slightly differing patterns. We have suspected this from the paper chromatograms, but have not yet investigated it properly.

Another interesting condition is that of Wilson's disease (hepatolenticular degeneration) (see also Uzman and Denny-Brown, 1949). The
shown as Plate XX,
acids into the urine,
Fanconi syndrome. Note the presence here too of an excessive cystine excretion. It would be ridiculous to talk of this as a 'cystinuria' but this mistake would be made if our method of analysis detected cystine only and not the other amino-acids at the same time. Moore and Stein (1952) have also analysed for many amino-acids the urines from five cases of Wilson's disease. They found the most remarkable resemblances in the amino-acid patterns of all the five cases, this strongly suggesting that the pattern is specific for the disease in question, as it is for fewer amino-acids in the case of the stone-forming cystinurics.

Now finally on Plate XXI, Fig. 16 is a urine from a new ob-

is characterized by skin changes and neurological signs. Again there is a gross amino-acid excretion, quite different from those previously shown and occurring constantly and in the presence

Let us turn back now to the renal amino-acidurias and other pattern abnormalities of lifelong origin. It is obvious that these must be hereditary and if so are likely to run in families. We have, therefore, with the collaboration of Dr. H. Harris, been studying this further. The work has already had some interesting results. The first concerns the method to be used in obtaining family data. It soon appeared that the urine analysis could be much more specific and unambiguous than the clinical syndrome. Harris therefore decided to restrict his first studies to

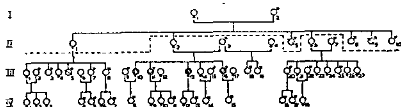


FIG. 17 Pedigree of the family of a case of Fanconi syndrome. The proposita is marked as a black circle Other cases discovered subsequently are shown as hatched circles.

urine testing of as many relatives as possible of known cases of renal amino-aciduria. Fig. 17 shows the unexpected result of his examination three years ago of the family of a case of Fanconi syndrome. The presenting case is shown in black. She had gross amino-aciduria and also very severe symptoms of osteomalacia and renal damage. None of her relatives gave any suggestive history of a similar disease; however the urine chromatograms showed that her two brothers and one of her two sisters had amino-aciduria. When they were examined clinically they were in good general health but blood and urine analyses showed all the expected signs of the Fanconi syndrome. They were therefore supposed to be in course of developing the disease Naturally they are being followed with great interest. One has already developed osteomalacia and is having to be treated. One has died of cancer and has provided us with valuable specimens, the third, the youngest in the family, is still well and is not being treated. It seems likely that the presenting case was severely affected with the bone disease years before her siblings because of the calcium depleting effect of her two pregnancies and lactations.

phenylketonuria or Folling's disease. Here the liver cannot metabolize the dietary phenylalanine which therefore accumulates in the blood and cerebrospinal fluid and is excreted in large quantities in the urine. In various liver diseases there are also disturbances of blood levels, especially of the blood cystine level (Dent and Walshe, 1951). The blood changes usually lead

TABLE 2. Classification of the Amino-acidurias showing their Subdivision into Two Types.

Mechanisms of Amino-aciduria

1. C [illegible]

I.e. [illegible] *presumably as a result of,* *rais* [illegible]

[illegible]

(c) Inborn, as in pheny[illegible] bolize dietary phenylalanine

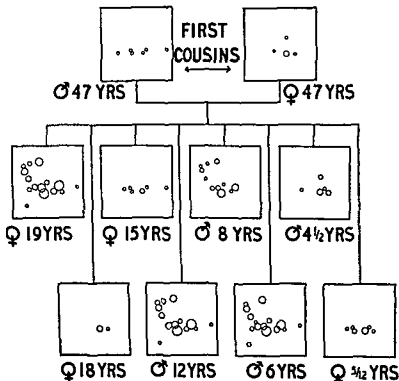
2. RENAL MECHANISM

I.e. Urine concentration increased while blood level remains normal. This usually involves different amino-acids to different extents so that the urinary amino-acid pattern is very variable from case to case, is constant in a given individual, and is quite unlike the blood pattern

(a) Acquired, as after lead and uranium poisoning (and presumably after other heavy metal tubule poisons also)
(b) Inborn, as in the Fanconi syndrome
(c) Inborn, as in Cystinuria
(d) Inborn, as in Hepato-lenticular degeneration (Wilson's disease)
(e) Inborn, as in several variants of the urinary pattern of amino-acid excretion, which occur in apparently normal people
(f) Inborn, as in 'Hart's syndrome'

to urinary changes ranging from a mild cystinuria to the gross amino-aciduria of acute yellow atrophy of the liver. These changes return to normal if good clinical recovery takes place. It is necessary to consider carefully the mechanisms likely to be causing excessive amino-acid excretion in any particular case. When it is the result of kidney tubule damage I like to refer to the phenomenon as a 'renal amino-aciduria', when the result of liver damage or other cause of a raised blood level, I call it an 'overflow amino-aciduria' These facts are summarized in Tables 1 and 2.

Now I shall revert to the interesting new syndrome ('Hart's syndrome') whose family Harris has been investigating. Fig. 19 shows the urine chromatograms in diagrammatic form arranged as a family tree. The mother and father were first cousins but



had normal urine. Of their eight children four have normal urine and four have an identical grossly abnormal renal amino-aciduria The clinical signs of these four are most confusing but two have both the neurological and skin changes while the two others so far have only the skin disease. Here again the urine changes are easier to identify than the clinical syndrome,

There is a further point of interest to be noted here. The syndrome has bred true. Each sibling had the same amino-acid pattern and associated signs. It is therefore probably a clinical individual. It also appears to have been inherited as a Mendelian recessive.

We have also shown that the pattern of the cystine-stone formers runs in families. Again many cases exist with the typical

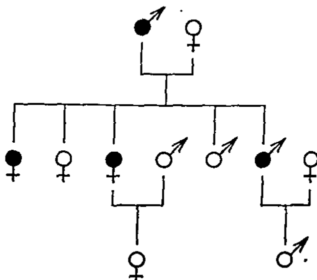


FIG. 18. Family tree showing as black circles the members who excrete unusually large quantities of β-amino-isobutyric acid ('T-spot').

pattern but no stone formation. Again this pattern breeds true. This is the final argument for the definite division into two diseases of this condition and the Fanconi syndrome. We can now separate them on the basis of genetic as well as of clinical and biochemical evidence.

Fig. 18 shows a family of normal people in which the pattern frequently occurs which is characterized by the high excretion of β-amino-isobutyric acid. There are four members with the pattern. That it is congenital and not environmental in origin is suggested by the fact that at the time of collection of the urine, one of the four was living in London, one was on a warship off the north of Scotland and two were living in Canada.

present to some extent in all body cells and is expressed as some abnormality in the capacity of the cell for taking up and concentrating amino-acids from its extracellular fluid. In the case of the kidney tubule we can detect such a defect easily from examination of the urine. In the case of other cells we can only analyse the blood. The blood levels are more difficult to interpret however as they express the results of the simultaneous operations of many factors acting in opposite directions. If this suggestion is correct it leads us on further. When a renal amino-aciduria has been found we must look for defects of some kind in other tissues too. In certain cases these are already known. For instance, in most children with the de Toni-Fanconi syndrome cystine deposits occur widely in liver, spleen, lymph, glands, etc. In Wilson's disease the liver contains large quantities of copper and, as Scheinberg and Gitlin (1952) have shown recently, the plasma proteins are lacking in caeruloplasmin, the protein fraction that normally carries in complex form the circulating copper. These more generalized defects have been discovered in various accidental ways. In my view we should systematically and directly study the possibility that other tissues are behaving like the renal tubule. This disordered behaviour of other tissues, if found, may be more easy to relate to the pathological condition of the brain, liver, bones and skin that occurs sometimes in association with renal amino-aciduria.

Another way of testing this hypothesis concerns a nutritional problem. It is well established that animals differ in their requirements of individual amino-acids. Sometimes this can be explained on the basis of known facts, as, for instance, in the case of the high cystine requirement of a good wool-bearing sheep where the cystine is obviously required in large quantity to build up wool protein. At other times, as in X-ray bacterial mutants, it appears to be purely genetic in origin. It is perhaps a reasonable speculation that an animal with a defective cellular mechanism for concentrating, say, methionine, from the extracellular fluid will demonstrate an increased dietary requirement for methionine. If this defect were also shared by the renal tubule it would reveal itself as a 'renal methionine-uria'. There would therefore be a correlation between the amino-acid

presumably because they are much more closely linked with the action of the abnormal gene.

General Conclusions

It is time now to look back on these preliminary results and see whether they lead us to any general conclusions, and whether we can learn from them any useful lessons which might help us in directing our future investigations.

The first clear fact is a rather practical one. It is that in a few rare diseases chromatographic investigations of urinary amino-acid excretions have a definite useful place as an aid to diagnosis. The second fact is a surprising one that so far defeats me for an explanation. It is that chronic renal amino-aciduria, a rare and spectacular finding, provides a link between diseases such as the Fanconi syndrome, Wilson's disease and I am sure many others, that do not resemble one another clinically, or in any other known way. Is our surprise at this due to the fact that when we try and think of the possible aetiology of a disease, we tend to overemphasize the value of our clinical findings and neglect to think enough of the ultimate underlying mechanism which must always lie, at least with congenital diseases, in the realm of atoms and molecules? The third fact is that in considering these renal amino-acidurias we are after all considering the varying functional capacity of the renal tubule cells, since it is almost unthinkable that the pattern changes can be explained other than by the presence of varying tubular reabsorptive capacities. Why however do we interest ourselves so much in the renal tubule? Nobody had thought till very recently that it was concerned in the pathology of Wilson's disease. Here I think we can put forward some likely speculations We must first reject the idea that the urinary losses of amino-acids produce an amino-acid deficiency. In no case so far is the loss equivalent to more than about 20 per cent of one's likely protein intake. My own present view is therefore that we have recently become interested in the renal tubule not because it is more important than other tissues nor because it is the actual cause of the disease, but rather because of the accident that it is more accessible to investigation. I like to think that the disordered gene action is

XIII

The Physiology of Parturition

A. ST. G. HUGGETT

THE cardinal point in parturition is the contraction of the uterine muscle, resulting in enlargement of the birth canal and leading to expulsion of the child and placenta. The study of parturition cannot be entirely divorced from the study of the changes undergone by the uterus during gestation and alteration from the non-pregnant state.

Following conception the uterus increases in weight from 50 gm. to 1,000 gm., in capacity from 5 ml. to 5,000 ml. and its muscle fibres grow in length individually from lengths of 50μ to lengths of 200–600μ. That is, a single nucleated cell grows in nine months to more than half a millimetre and becomes visible to the naked eye. This is accompanied by nitrogen retention and deposition dependent upon secretion of the anterior lobe of the pituitary and overaction of the islets of Langerhans (Young, 1944, 1948). The position was considered more fully in a discussion at the Royal Society of Medicine (Huggett, 1951) following upon a paper by Dr. S. R. M. Reynolds (Reynolds, 1951).

In this lecture it is proposed to outline the normal orthodox views on the uterine muscle action as seen by the obstetric physiologist (or is it the physiological obstetrician?). Then one will consider the ways in which the movements of the uterus in parturition can be studied, then how they are modified by hormones and drugs, discussing their responsibility in producing these changes. Lastly one will consider the application to the problem of certain modern views on biophysics and biochemistry.

requirement of an animal and its urinary amino-acid pattern, both being readily determinable experimentally.

I hope I have succeeded in illustrating how new techniques of analysis have led to new knowledge of disease processes, and how, in the amino-acid field in particular, great advances are likely to be made in the near future.[1]

[1] Since the text of this lecture was prepared for press further papers dealing in more detail with some of the subjects discussed here have appeared as follows:
'Family Studies on the Urinary Excretion of β-Aminoisobutyric acid.' H. Harris *Ann. Eugenics*, 1953, 18, 43
'Urinary Amino-acid Patterns of some Mammals.' S. P. Datta and H. Harris. *Ann. Eugenics*, 1953, 18, 107
'Quantitative Studies on the Urinary Cystine in Patients with Cystine Stone Formation and in their Relatives.' H. Harris and F. L. Warren. *Ann. Eugenics*, 1953, 18, 125.

REFERENCES

Consden, R., Gordon, A. H. and Martin, A. J. P. (1944). *Biochem. J.* 38, 224.

Crumpler, R. and Dent, C. E. (1949). *Nature*, **164**, 441.

Crumpler, R., Dent, C. E, Harris, H. and Westall, R. G. (1951). [illegible]

[illegible] oundation Symposium, London

Moore, S and Stein, W. H (1951). *J. Biol Chem.* **192**, 663.

Moore, S and Stein, W. H (1952). *2nd Internat. Congr. Biochem. Abstracts*, p. 367.

Partridge, S. M. and Westall, R. G. (1949). *Biochem J.* **44**, 418

Scheinberg, I. H. and Gitlin, D (1952) *Science*, **116**, 484.

Stein, W. H. (1951). *Proc. Soc exp Biol. and Med.* **78**, 705.

Uzman, L and Denny-Brown, D. E. (1949). *Amer J Med Sci.* **215**, 599.

Westall, R. G (1952). *Biochem. J.* **52**, 638.

METHODS OF STUDY OF UTERINE CONTRACTIONS

The study of uterine contractions can be carried out in a number of ways.

(i) By clinical observation using clinical methods, including not only palpation but also radiological techniques, including in animals silver wire on uterine contours (Bell, 1941).

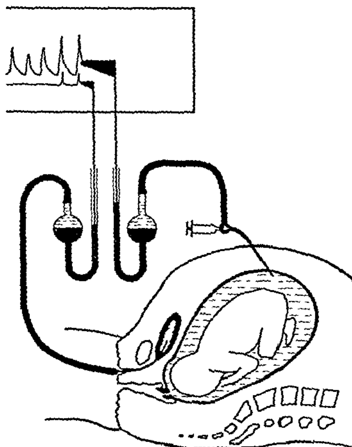


FIG. 1. Recording the intrauterine and intraabdominal pressures with cannula in the amniotic sac and vesical bladder respectively (after Alvarez and Caldeyro, 1950)

The Stages of Labour

The term 'birth canal' was first applied by the late John Fairbairn to the passage along which the foetus proceeds to the exterior. He pointed out that the uterus did work in regard to the birth canal in two respects: (i) by enlarging the canal to an adequate diameter for the foetus to reach the exterior, and (ii) by reason of the species difficulties presented by its curvature in the human.

The birth canal is divisible into two parts, the upper and lower uterine segments. The upper uterine segment is that part of the uterus entirely covered by peritoneum. The lower is the portion of the uterus which is not so covered, together with the cervix and vagina. In the upper uterine segment the fibres run in all directions. In the lower they are longitudinal and circular mainly, but absent for all practical purposes in the vagina. The upper uterine segment hardens in labour by contracting, and stretches the cervix and vagina, pulling them up. The lower uterine segment thereby becomes thinned out and lengthened and the wedge effect of the membrane dilates the cervical canal.

The placenta and membranes crenate off and loosen from the retracting muscle, beginning even before foetal expulsion. For didactic purposes three stages of labour are defined separately, though myometrically they merge one with the other. They are defined as

Stage 1. From the commencement to full dilatation of the cervix;

Stage 2. To expulsion of the child;

Stage 3. To expulsion of the placenta.

In the first stage not only is the contracting, thickening upper uterine segment pulling up and thinning out the cervical canal, including the lower uterine segment, but the intrauterine pressure is rising and the membranes are pushing down the canal, as a bag of waters, acting as a wedge dilating the cervix. At the same time there is a softening of the cervix which helps the dilating action of both the membraneous wedge and pulling upper uterine segment.

NORMAL MOVEMENTS

The uterus exhibits throughout pregnancy mild rhythmic contractions which become more intense as gestation proceeds (Fig. 3). Using intraamniotic pressure recordings, Alvarez and

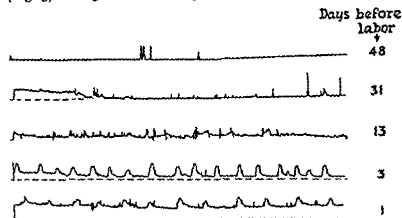


FIG. 3. Prelabour uterine motility at varying periods before the onset of labour (Murphy, 1940).

Caldeyro (1950) have demonstrated that these are of two types, (i) rhythmic contractions (10–30 per 10 minutes) of low intensity, developing pressures under 8 cm. of water (6 mm. Hg.), and (ii) occasional high-pressure contractions exceeding 8 cm. of

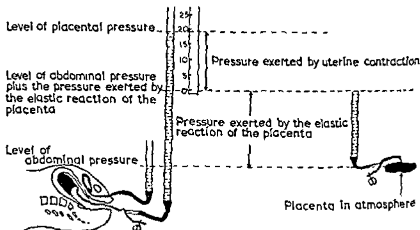
TABLE 1. Frequency (contractions per 10 min.), 'tonus' and 'intensity' (measured in cm. of water pressure)

	Frequency	Tonus	Intensity
Pregnancy—Arhythmic	Very rare	—	8 cm.
—Rhythmic	10–30	4–11 cm.	0 3–7 cm.
—Last 2 weeks	3–9	8–15	6–35
Labour—1st & 2nd stages	2–6	8–17	25–70
—3rd stage	2–5	5–10	35–60

water—the 'painless contraction' of pregnancy or Braxton-Hicks type. They increase in frequency in the last fortnight of pregnancy (Fig. 4).

(ii) By recording externally the shape and consistency of the uterus relative to the abdominal wall, using a tocograph (a recorder of the anaeroid barometer type) strapped on the body wall. A modification, the tocodynamometer, is a type of strain gauge.

(iii) By registering the movements of the uterus *in vivo*. This can be done in animals or in humans, in the case of the latter balloons being inserted to record changes of intrauterine pressure.



[illegible] pressure exerted by the uterine contraction (after Alvarez and Caldeyro, 1950).

(iv) By reading the intrauterine pressure in the human by (*a*) recording the intraamniotic pressure with direct puncture (Fig. 1), (*b*) recording the intraplacental or intraumbilical vein pressure (Fig. 2). (Alvarez and Caldeyro, 1950).

(v) By perfusing the uterus and recording its movements (Kurdinovsky, 1906; Kehrer, 1907).

(vi) By registering the movements *in vitro* of isolated muscle strips.

(vii) By isolating the uterine muscle proteins and recording the elastic and contractile powers of the purified proteins from different uteri in different metabolic conditions (Csapo, 1948).

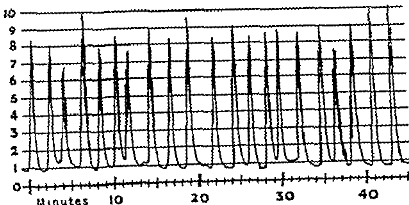


FIG. 5. Uterine movements in second stage of labour Each uterine contraction has superimposed smaller waves denoting bearing down efforts by the patient (Alvarez and Caldeyro, 1940).

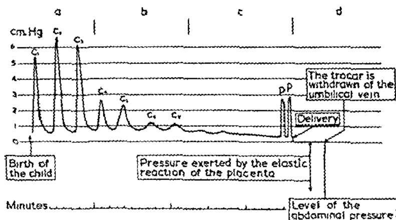


FIG. 6 Placental pressure in the third stage of labour.

a. While in the upper uterine segment.
b While leaving the upper uterine segment.
c While in the lower uterine segment and birth canal
d After the third stage and after placental delivery

When labour supervenes these contractions increase in intensity to 4–9 cm. Mercury pressure and a rate of 2–9 per 10 minutes. They continue during the three stages of pregnancy but peter out as the placenta leaves the uterus (Figs. 5 and 6).

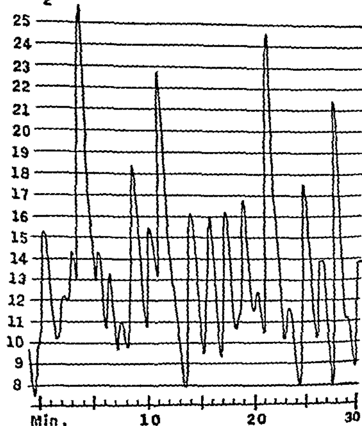


FIG. 4. Uterine movements Tracing in cms. pressure of water recorded 2 days before labour by Alvarez and Caldeyro (1950).

Table 1 shows how these pains can vary throughout gestation and labour.

The effect of 'bearing down' upon these contraction records is the creation of spikes.

Reynolds, Heards and Bruns (1947, 1948) have been responsible for the use of the tocodynamometer. This consists essentially

Reynolds was able to calculate the tensile strength of the cervix, the rupturing force of the membranes, and to contrast the action of the membranes before and after rupture (Reynolds, 1949).

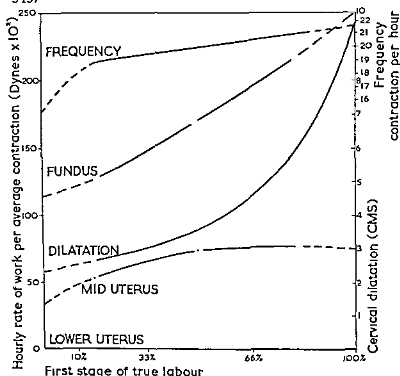


FIG 8 Uterus, first stage of labour. The frequency of fundus contraction, the cervical dilatation in centimetres and the work in dynes per hour of the fundus, mid uterus and lower uterine segment throughout labour (Reynolds, 1951)

Combining forces with Alvarez and Caldeyro in Montevideo, Reynolds used his tocodynamometer on the same cases in which Alvarez and Caldeyro punctured the amnion (Caldeyro, Alvarez and Reynolds, 1950).

In consequence of their teamwork they were able to show that the clinically 'good labours' were those in which the different parts of fundus and mid-uterine zone contracted with good

of a strain gauge strapped on to the abdominal wall and capable of recording local contractions of the uterus by the pressure exerted upon the wall at the point of strapping of the strain gauge. With this instrument Reynolds and his colleagues draw the surface markings of the uterus upon the abdomen, and have recorded the local changes at several different points (up to 7) on the uterus: for this Reynolds uses a 7-channel recording for 7 tocodynamometers. He is therefore able to record differences

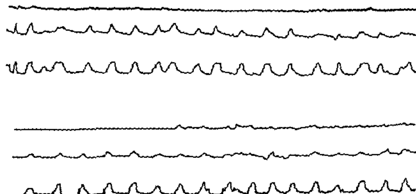


FIG 7. Record of uterine movements with the tocodynamometer during the 8th and 9th hours of labour Each trio of lines represents in order from above down the simultaneous activity in the lower uterine segment, mid portion and the fundus of the uterus respectively (Reynolds, 1949).

in contractility in different parts of the uterus or changes produced by different treatments. The contractions of the fundus, mid-uterine zone and lower uterine segment were recorded, and it was found that in the first stage up to dilation of the cervix, the fundus contractions are intermittent but strong, and rise to maximum strength quickly and have a relatively long duration. Those in the mid-zone are shorter and diminish progressively in intensity and continuance with labour continuance while the lower uterus segment is inactive (Fig. 7). In consequence of this, Reynolds and his colleagues were able to make an assessment of the relative amounts of work performed by different parts of the uterus during labour. Fig. 8 shows this. The lower uterine segment does no work during labour and the fundus most. Its increased work with greater frequency of contractions effectively dilates the cervix and lower uterine segment.

muscle is 5 mg. per gramme of muscle, but it rises to 13 mg. at full term. At the same time myosin decreases from 11 to 9 mg. per gramme. This is best shown in table form.

TABLE 2. Amounts of Myosin and Actomyosin in Human Muscle mg./g.

	Myosin	Actomyosin
Striped muscle	18	26
Non-pregnant uterus muscle	11	5
Pregnant uterus muscle	9	13

Csapo (1950b) has also shown that ovariectomy in rabbits lowers the actomyosin content of the uterus from 8 mg./g. of uterus to less than 2 mg./g. and the total content from 35 to 7 mg.

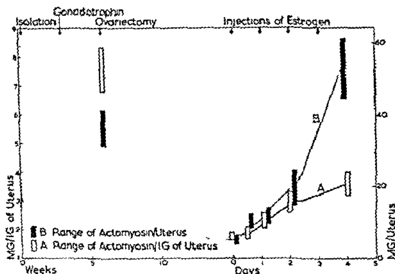


FIG 9 Rabbit's uteri. The concentration of actomyosin in mg. per gramme of uterus and mg per uterus together with the effects of ovariectomy and of subsequent oestrogen injections (Csapo, 1950b).

When oestrogens were given to these ovariectomized rabbits, the actomyosin content increased (Fig. 9).

A curious third member of the triumvirate is adenosine triphosphatase. Myosin can be got crystalline. It is, however, still a

synchrony. Asynchrony occasionally occurred. A good rise in amniotic pressure occurred with synchrony and high intensity of contraction—it would rise by more than 24 cm. of water. Further, a good labour would be characterized by regular rhythm and good relaxation with full descent of the amniotic pressure.

Causation and Control of Labour

Labour can occur spontaneously in cases with extrinsic nerves cut or with the spinal cord damaged in the region of the lumbar centres (Routh, 1897, Kurdinovsky, 1904). The unloaded uterus shows rhythmic contraction (Bell, 1941, guinea-pig). Spontaneous activity appears to increase during oestrus. In the first part of pregnancy these movements disappear, but at the ninth week (human) they return (Knaus, 1927; Alvarez and Caldeyro, 1950). In the animal, oestrogens appear to increase activity of the uterus and also the response to oxytocin which alone has a general effect in increasing uterine contractions, an effect in humans which appears to increase with the approach of parturition (Robson, 1933). The effect of progesterone is not clearly antagonistic to that of oestrogens. It inhibits the rabbit's uterus provided this has been sensitized with oestrogens (Robson, 1932). On the other hand, in man there is some evidence of slowing of movement but not of decrease of amplitude (Bell, 1941).

The Contractile Substance of Uterine Muscle

In 194 [illegible]
could [illegible]
B visci [illegible]
third p[illegible]
colleague Straub (1942, 1943). This new myosin *B* formed with actin, Szent-Gyorgi called actomyosin.

It appears to be the structural protein of the resting muscle fibrils, and is itself composed of a fibrous network (Astbury, Perry, Reed and Spark, 1947). One gramme of moist striped muscle contains about 30 mg. of actin, 70 mg. of myosin A and 0·2 cc. of water. Csapo (1948, 1950a), a pupil of Szent-Gyorgi, has shown that the actomyosin content of non-pregnant human

Myosin can be digested by trypsin to small molecules (Perry, 1951). The mixture so obtained retains its ATP-ase or enzyme activity, but has lost the power of forming actomyosin threads or increasing in viscosity on mixing with actin.

In 1930 Lundsgaard showed that muscle poisoned with iodoacetic acid continued to contract *in vitro* in nitrogen on stimulation but did not exhibit glycolysis and formed no lactic acid. In 1934 Lundsgaard demonstrated this to be accompanied by a loss of adenosine triphosphatase (ATP). The number of contractions obtainable was about 100 and exhaustion synchronized with loss of the ATP stores. Lohmann (1934) simultaneously transferred phosphate from creatine phosphate to adenosine diphosphate to form the triphosphate. It is therefore now considered that this complex of creatine phosphate plus ATP forms a store or reserve of energy, utilizable for anaerobic contraction of muscle capable of exhaustion unless replenished aerobically. This energy reserve is within the high-energy phosphorus bonds of the muscle. The conversion of glucose to glycogen *en route* to triose phosphate and oxidation to CO_2 and water liberates energy in the process which is available for reformation of creatine phosphate and ATP. This liberation of energy is an aerobic process and requires not only oxygen but potassium

This raises the question of its application to uterine muscle, a problem which has been prosecuted in Baltimore by Csapo and in Norway by Walaas.

Creatine phosphate and ATP are present in uterine muscle, but only to one-sixth to one-tenth of that in skeletal muscle, together being 2 64µM./g. concentration (Caspo and Gergeley, 1950; Walaas and Walaas, 1950a). Ovariectomy reduces this concentration in rats and oestrogen injections increase it and also cause an increased turnover of total acid-soluble phosphorus compounds as adjudged by the incorporation of radioactive phosphorus (Walaas and Walaas, 1950b).

Csapo and Gergeley (1950) have determined the number of anaerobic contractions to exhaustion of idioacetic acid-treated uterine muscles from rabbits. This is about 10, approximately one-seventh to one-tenth that of skeletal muscle (psoas from rabbit). Dividing this into the concentration of ATP and creatine

mixture containing muscle enzymes, notably adenosine triphosphatase (ATP). This myosin Szent-Gyorgi *claims* will hydrolyse the ester ATP itself. But on the other hand the addition of ATP to actomyosin will cause it to lose its viscosity. At the same time *there* is a loss of flow birefringence. (Birefringence together with

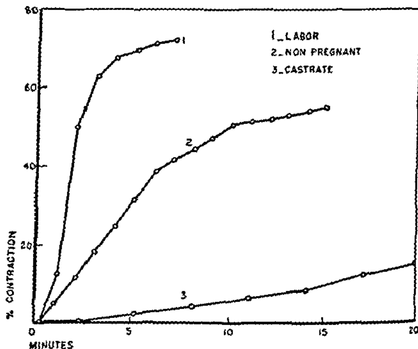


Fig. 10. Contractile power, measured as percentages of the maximum contraction over a period of time subsequent to excitation, for actomyosin threads obtained from 1, rabbit uterus in labour, 2, non-pregnant rabbit uterus, 3, castrate rabbit uterus (Csapo, 1948).

viscosity is an index of long thread-like molecules capable of contraction.) The electron microscope has demonstrated (Perry and Reed, 1947) that the actomyosin molecules have gone to globular forms which will not contract. Now Csapo has shown (1948) that actomyosin threads from human uteri can be made to contract and the contractile power varies with samples obtained *in* different states of the uterus (Fig. 10), being maximal in those from uteri in *labour* and minimal from uteri of castrates.

REFERENCES

ALVAREZ, H. and CALDEYRO, R. (1950). *Surg. Gynec. Obstet.* 91, 1.

ASTBURY, W. T., PERRY, S. V., REED, R. and SPARK, L. C. (1947). *Biochim. biophys. Acta*, 1, 379.

BELL, G. H (1941). *J. Physiol.* 100, 263.

CALDEYRO, R., ALVAREZ, H and REYNOLDS, S. R. M. (1950). *Surg. Gynec. Obstet.* 91, 641.

CORNER, G. W. and CSAPO, A. (1952). Ciba Foundation Lecture. 1952.

11, 701.

Bull Johns Hopkins Hosp. 82, 466.

REYNOLDS, S R. M. (1949). *Physiology of the Uterus.* 2nd edn. Hoeber, New York

r. obstet. Soc. Lond. 29, 191.

2, 3: 3, 23.

phosphate yields 2·64/10, that is 0·264. This value of about 0·25μM. of ATP plus creatine phosphate per gramme of muscle per contraction is identical with that obtained by A. V. Hill (1949, 1950) for skeletal muscle via the direct measurement of the initial heat of a single striated muscle contraction (3m. cal.) from which he determined the ratio of the high-energy phosphorus per gramme of contraction as the above value.

Corner and Csapo (1952) have studied the tension of uterine muscle on stimulation when suspended in Kreb's solution. They found that the maximal tension developed varied with frequency of stimulation. If the uterus was oestrogen dominated (treated before death and removal with oestrogen), decreasing frequency resulted in decreasing tension, and if progesterone dominated, in increasing tension in a staircase effect. They correlated these effects with the concentration of potassium inside the muscle and its rate of removal from the interior to the exterior. Their key point is that potassium is essential for the aerobic glycolysis, and liberation of energy for the replenishment of the high-phosphate bond energy of the muscle.

Summary

this problem. Omission has also been made of the nerve paths involved, the mechanism of action and interaction of endocrines and local hormones upon the muscle fibre, the variations in blood flow in different parts of the uterus, and of the changes occurring in hormone balance in the blood before and at the onset of parturition, and attention has been confined to three new methods of study and the big step that has been taken in the study of the properties of the actual contractile proteins involved.

because all metabolic processes would be arrested or reduced to a negligible level at such temperatures. So far as vertebrate cells are concerned there had been little opportunity of investigating this idea because with few exceptions the processes of freezing and thawing had been found to be lethal to the cells of higher animals. In the course of our work, however, we made the discovery that glycerol has remarkable properties in protecting certain spermatozoa against the effects of freezing and thawing, and we have been able to extend this discovery to other cells and to certain tissues. As a result, we have found out much about the behaviour of living cells during freezing and thawing and during maintenance at low temperatures.

I must not give the impression that the freezing and thawing of living cells is a comparatively new field of experimentation. There is an extensive literature on the freezing of invertebrate organisms, plant cells and tissues, and unicellular organisms. A large amount of work has also been done on the freezing of tumour cells in vertebrates. So far as normal vertebrate cells are concerned, however, and I refer in this lecture only to normal cells, the literature is comparatively scanty.

Spermatozoa

I have indicated that the starting point of our work was the freezing and thawing of spermatozoa. One of the earliest observations, made by Davenport in 1897, was that a proportion of human spermatozoa revive after freezing and thawing untreated semen by comparatively crude methods, and this observation has been amply confirmed (see Parkes, 1945). Unfortunately, the human spermatozoon appears to be unique in this connection, and little progress was made with other types of vertebrate gametes. Luyet and Hodapp (1938) had shown that spermatozoa of the frog could be revived to some extent after freezing and thawing in strong sugar solutions, and more extensive results on the same lines were obtained by Shaffner, Henderson and Card (1941) with fowl spermatozoa. The results obtained by these methods, however, were poor, and the few eggs fertilized by the frozen and thawed fowl spermatozoa failed to develop (Shaffner, 1942).

XIV

Preservation of Living Cells at Low Temperatures

A. S. PARKES

Introduction

It is not my intention to give a review of the whole field covered by the title of this lecture. I want to tell you something of the work we have been doing at the National Institute for Medical Research in the last four years, on the freezing and thawing of living cells, mainly of mammals, and on their storage at low temperatures. I should add that my colleagues Dr. Audrey Smith and Mr. Polge, and later Dr. Lovelock, Mr. S. E. Smith and Dr. Deanesly, have been intimately concerned with the work I have to talk about. I want to describe this work as an example of practical day-to-day research rather than as the pursuit of a logical concept, because it provides a very good example of how one thing leads to another and of the lack of any clear distinction between so-called fundamental and so-called applied research.

Some years ago in the course of work on the effects of physical and chemical agents on the gametes we became interested in the effects of low temperatures on spermatozoa, and when I speak of low temperatures I refer to what biologically are very low temperatures such as can be obtained by the use of solid carbon dioxide at $-79°$ C. or by the use of liquid air at approximately $-190°$ C. In approaching this subject we had in mind the idea which, of course, was by no means original, that if a living cell could be frozen and thawed again without damage then the interval between these two processes should not be important

problem of freezing mammalian spermatozoa. The spermatozoa of the rabbit, the only laboratory mammal convenient for this type of work, have proved to be extraordinarily sensitive to freezing, but results of great theoretical interest and practical importance have been obtained with bull spermatozoa. These results were obtained only after long experimentation and depend on two major points of difference from the technique used for fowl spermatozoa—the prolonged equilibration of the

TABLE 1. Long-term Storage of Bull Semen at −79° C. (Dilution 1 : 4)
(Taken from Polge and Rowson, 1952b)

Age of semen (weeks)	No. of cows	Conception rate* (per cent)
1–5	16	63
5–8	16	69
9–12	16	81
13–16	16	63
17–20	16	50
21–24	16	81
25–28	16	56
29–32	16	75
	Average conception rate	69

* Half the cows in the groups up to 28 weeks have been examined for pregnancy. The remaining cows have not returned to oestrus within three months from insemination. In the group 29 to 32 weeks the conception rate figure represents non-returns at two months

spermatozoa with the glycerol-containing medium (Polge and Rowson, 1952a) and slow cooling (Smith and Polge, 1950b). With these elaborations of technique it has been possible to freeze and thaw bull spermatozoa without apparently affecting their capacity to fertilize eggs after insemination. Moreover, long-term experiments have so far failed to show any falling off in material maintained at −79° C. (Polge and Rowson, 1952b) for up to eight months (Table 1). There is at present the complication that the concentration of spermatozoa used has been greater than is normally necessary, and storage experiments on higher dilutions than 1 : 4 are in their early stages. Subject to this qualification it seems that bull spermatozoa, under the conditions used by us, maintain a stable state at −79° C. These

In attempting to repeat the work of Shaffner, Henderson and Card we failed to obtain even their modest degree of success, but during our attempts we made the discovery, to which I have already referred, about the remarkable properties of glycerol (Polge, Smith and Parkes, 1949). The effect of using diluents containing glycerol is spectacular, as shown by the following simple experiment. A sample of fowl semen is obtained and divided into two parts. One part is diluted with an equal volume of Ringer's solution, and the other with an equal volume of Ringer's solution containing 30 per cent glycerol. The two specimens are frozen by plunging into alcohol cooled to −79° C. with solid carbon dioxide, and subsequently thawed in a water bath at 40° C. On examination it is seen that the spermatozoa in the sample diluted with Ringer's solution are all dead; those in the sample diluted with the glycerol-containing diluent are indistinguishable from normal and are intensely active, giving to the specimen the swirling appearance typical of fowl semen. Early experiments on the fertilizing capacity of such frozen and thawed material were disappointing (Smith and Polge, 1950a), but it was then found that the presence of glycerol with or without freezing was inhibitory to fertilization and that frozen and thawed samples from which the glycerol had been removed retained normal fertilizing capacity (Polge, 1951). With this information available experiments on long-term storage were begun, and it was found that at −79° C. fowl spermatozoa deteriorated rapidly so that, after a few weeks, motility on thawing, and with it fertilizing capacity, was very poor. At −190° C., however, survival was much better and normal chicks have been obtained from spermatozoa stored for as long as nine months at this temperature (Polge and Parkes, 1952). Even at −190° C. there is a slow but steady loss of spermatozoa. Nevertheless, those that survive appear to be perfectly normal, and in the course of our work two generations of birds have been produced from a single sample of semen. These results with fowl spermatozoa provided the first indication of a most important fact, that even at very low temperatures the cell is not necessarily in a stable state.

Against this background we made another approach to the

the circulation, and there have a normal life-span. In other words, the surviving cells have not aged metabolically during storage. The comparative survival of frozen and unfrozen red cells reintroduced simultaneously into the circulation is shown in Fig. 1 taken from the paper by Mollison and Sloviter (1951).

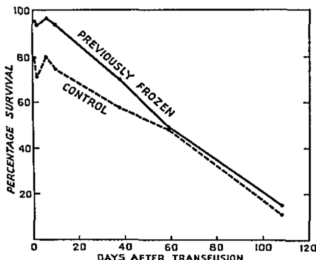


FIG 1. The percentage survival of previously frozen red cells and control red cells at intervals after transfusion. (Fig. 2 from the paper by Mollison and Sloviter, 1951, which should be consulted for details. *Reproduced by permission of The Lancet Ltd.*)

FERTILIZED OVA

After the early successes in the freezing of spermatozoa attention was turned to the fertilized egg of the rabbit. Initially the experiments were unsuccessful, but later it was found that the fertilized but unsegmented egg of the rabbit could be treated with glycerol, provided it was applied gradually and at body temperature, and that some eggs so treated would begin segmentation in culture after exposure to liquid air (Smith, 1952a). The survival rate was very low, little more than 1 per cent, but it was enough to demonstrate that even a cell in such an unstable condition as a newly fertilized egg does not necessarily succumb on exposure to low temperatures (Plate XXII, Fig. 2) [1]

[1] The plates referred to in this lecture will be found at the end of the book.

results with bull spermatozoa are now being widely applied to the technique of breeding cattle by artificial insemination and promise to revolutionize erstwhile practice.

Red Blood-cells

The work on spermatozoa has expanded in various directions, notably to the red blood-cell. During the work with the glycerol-containing semen diluents Dr. Smith became interested in the osmotic properties of such solutions and used red blood-cells as convenient indicators. Rather surprisingly, she found that rabbit red cells behaved quite normally in semen diluents containing 15 per cent glycerol. The effect of freezing under these conditions was then investigated, and it was found that the glycerol-containing diluents protected red cells against haemolysis, as they protect fowl and bull spermatozoa against the otherwise fatal effects of freezing and thawing (Smith, 1950). Confirmation of these observations was readily forthcoming (Sloviter, 1951). In these early experiments similar results were also obtained after periods of freezing of up to three months.

On this foundation a considerable research on the preservation of human red blood-cells was built by Dr. H. A. Sloviter, of Philadelphia, then working at the National Institute for Medical Research, and Dr. P. L. Mollison of the M.R.C. Blood Transfusion Research Unit, and it is now possible to keep whole blood in bulk (Mollison and Sloviter, 1951) for long periods (Mollison, Sloviter and Chaplin, 1952) by using glycerol diluents and low-temperature storage After the removal of the glycerol by dialysis (Sloviter, 1951) or washing (Lovelock, 1952) such blood can be used for transfusion. Dr Mollison in his lecture in this series on the life-span of the red blood-cell will no doubt be dealing in detail with this matter, and I need therefore do no more than call attention to one important point The loss of red cells by freezing and thawing of small samples by this technique is of the order of 10 per cent. On prolonged storage, however, there is a further small but continuous loss of cells, so that after six months recovery is down to about 70 per cent (Sloviter, 1952). With bulk preparations there is a rather greater loss, but the cells that survive behave normally when reintroduced into

is only about seven days. This test has proved remarkably good for investigating the effect of treatment of the ovarian tissue before grafting, and a quantitative estimate of the effects can be obtained by considering the percentage of active grafts obtained, and the average time taken to re-establish the vaginal cycle (Smith and Parkes, 1951).

TABLE 2. Results of Grafting Ovarian Tissue frozen for nine days in Various Media, in comparison with controls
(Taken from Smith and Parkes, 1951)

Treatment of ovarian tissue	Proportion of animals showing active graft within 28 days	Average time to establish active graft (days)
Unfrozen control:		
15% glycerol-saline at room temperature for one hour	14/14	6·9
Frozen in:		
15% glycerol-saline to −190° C.	5/5	14·4
Frozen in:		
15% glycerol-*serum to* −190° C.	5/5	16·4
Frozen in:		
25% glycerol-saline to −190° C.	3/5	23·3
Frozen in:		
Serum to −79° C.	4/4	24·0

Using this test and the technique of slow freezing worked out for the bull spermatozoa, we examined systematically combinations of conditions obtained by varying the medium (saline, serum and 15 per cent and 25 per cent glycerol in both), final temperature (−79° C. and −190° C.) and time of freezing (one hour, and nine days). Twenty-two of the possible 24 combinations were examined. Some of the treatments resulted in no ovarian grafts being produced. With others, and this is a very important point, conditions which permitted the survival of tissue frozen *for one hour did not permit of survival when frozen for nine days.* The best results for the nine-day period of freezing are shown in Table 2, from which it will be seen that the most effective grafts were obtained from tissue which had been frozen in 15 per cent glycerol-saline at −190° C. With this medium a high proportion of active grafts was obtained, but the average time taken

Endocrine Tissue

So far, however, the most interesting development of the egg work has been indirect. Early on it was noticed that although the egg itself failed to develop after freezing and thawing, the cumulus cells of the follicular granulosa still adhering to the egg after ovulation, grew in culture to some extent after the treatment (Smith, 1949). This result was confirmed in other experiments in which cumulus cells were separated from the egg, frozen, thawed and cultured after treatment by various methods (Smith, 1952b). Plate XXII, Fig. 3 shows cumulus cells of the rabbit proliferating in culture after having been frozen to −79° C. for 24 hours.

This observation on cells essentially endocrine in nature prompted an extension of the work to ovarian tissue in general and thence to other endocrine organs. At this stage, therefore, we had to review relevant literature. So far as the mammalian tissue is concerned it seems that there is one record of thyroid tissue having been successfully grafted after freezing (Blumenthal and Walsh, 1950), but the most impressive observations relate to mammalian skin, which has a comparatively high resistance to freezing (Medawar and Billingham, 1952). The use of freezing for the preservation of certain other tissues commonly grafted in surgical practice is, of course, well known, but in the case of artery and nerve it is likely that the treatment kills the tissue, which when grafted acts as a framework for the growth of new tissue from the host.

In any work of this kind it is of prime importance to arrive at a simple and unequivocal test for the survival of tissue after treatment. In the case of ovarian tissue this is particularly easy. If the ovaries are removed from an adult rat, chopped into pieces and reimplanted into the subcutaneous space of the flank, functional grafts are made very rapidly, as is shown by the restoration of the cyclic changes in the vaginal smear, which are interrupted by ovariectomy. This reaction occurs with remarkable regularity and remarkable speed. Under quite crude conditions autografts of this kind take in almost every animal, and the average time for the re-establishment of the vaginal cycle

conflict of opinion on this point which impinges on the wider problem of the reformation of the oocytes by the germinal epithelium of the adult. The fact is, however, that we have observed very few oocytes in frozen and thawed ovarian tissue for up to three days after grafting, and that thereafter more primordial oocytes seem to appear. At a later stage the graft contains a number of large follicles (Plate XXIII, Fig. 4), and corpora lutea which are atretic in the sense that they develop without ovulation, which is inhibited by the topography of the graft. It seems, however, that the graft of frozen tissue ultimately loses its capacity to form oocytes and such a graft eventually comes to consist purely of amorphic nodules of secretory tissue (Plate XXIII, Fig. 5) (Parkes and Smith, 1953). This fact is of no practical importance in the case of ovarian grafts made subcutaneously since even the cyclic endocrine activity of the graft may be maintained in the absence of cyclic structures (Parkes, 1952b). The preservation of the oocytes during the freezing and thawing of ovarian tissue must, however, receive attention before any attempt can be made to use low-temperature storage for tissue intended for orthotopic grafting and subsequent fertility of the host.

There is another way in which frozen and thawed tissue appears to differ from normal tissue. Fresh ovarian tissue of the rat can be kept at room temperature for several hours, or at 0° C. for a day or more, without losing its capacity to make a functional graft. The frozen and thawed tissue, by contrast, is much more labile. Tissue frozen in 15 per cent glycerol-saline and kept at room temperature after thawing loses almost entirely its capacity to form a functional graft in less than three hours, so that the percentage of takes is very much reduced. To try to determine the reason for this lack of viability in frozen and thawed tissue other experiments were carried out on freezing in glycerol-serum, and on maintaining the tissue after thawing at 37° C. or at 0° C. The full results are given in Table 4, from which it will be seen that the use of serum rather than saline gave better conservation of the tissue at room temperatures At 37° C., however, the tissue deteriorated rapidly in either medium, while at +2° C. conservation was fairly good with

was about twice that for normal tissue, indicating presumably that there had been a considerable loss of cells on freezing and thawing, and that the remainder had taken a comparatively long time to build up an effective graft. It is noteworthy that the 15 per cent glycerol-saline did not permit the survival of tissue for nine days at −79° C. The other media gave results less good and need not be further mentioned.

Having found conditions which appeared to be promising for the maintenance of ovarian tissue at low temperatures, we proceeded to a long-term experiment. Table 3 lists results for

TABLE 3 Retention of Viability in long-frozen Ovarian Tissue

Duration of freezing of tissue (days)	No. of implantations	No. of active grafts obtained	Average time taken for reappearance of cycle (days)
1/24	5	5	15 4
9	5	5	14 4
83	5	5	18 8
117	8	8	15 1
235	9	9	16 0
367	12	12	16 3

freezing ovarian tissue in 15 per cent glycerol at −190° C. for periods varying from one hour to one year. It will be seen that, as already mentioned, there is clearly some loss of cells caused by freezing and thawing, but the loss is not increased by extending very greatly the interval between these processes. It seems, therefore, from these experiments with ovarian tissue that the medium has a very significant effect upon the stability of the cell maintained at low temperatures over a long period.

All I have said so far about ovarian tissue relates to its endocrine activity and specifically to its oestrogenic activity. In one way the graft made from frozen tissue has differed essentially from that made from normal tissue. The process of freezing by present methods abolishes nearly all oocytes, and the occurrence of oocytes in the graft later on may perhaps depend partly upon the regeneration of oocytes, presumably from surviving germinal epithelium. I say perhaps, because there may well be

Steinach and of Lipschutz on the production of intersexuality by the grafting of the heterosexual gonad, and all similar work in mammals, has necessarily involved homografting. With these various considerations in mind a number of experiments on the homografting of frozen rat ovarian tissue were carried out. The results are summarized in Table 5 from which it will be seen

TABLE 5. Comparison of Autografting and Homografting of frozen Ovarian Tissue

Type of graft	Treatment of grafts	No. of implantations	No. of active grafts obtained	Average time before cycle re-established (days)
Autograft	Chopped in 15% glycerol-saline; soaked for one hour. Not frozen.	14	14	6 7
Homograft	Chopped in 15% glycerol-saline, soaked for one hour. Not frozen	25	18	10 9
Autograft	Chopped in 15% glycerol-saline; frozen to −190° C for 9 days	14	12	15 6
Homograft	Chopped in 15% glycerol-saline, frozen to −190° C. for 9 days.	20	9	21 9

(1) that the unfrozen homograft 'takes' freely, but not quite as well as the autograft, and the time taken to establish an active graft is longer; (2) that the percentage of 'takes' is again somewhat reduced, and the 'taking' time much increased, by the previous freezing and thawing of the tissue. We have not at present sufficient data for a fully quantitative evaluation of this problem, but it is already clear that the technique of preservation at low temperatures can be applied to tissue meant for homografting without any great addition to the difficulties inherent in that process.

both. Even the best of the results recorded in Table 4 are much inferior to those which are obtained under similar conditions with unfrozen tissue, and at present we are unable to offer any explanation of the rapid deterioration of frozen tissue after thawing. It is clearly essential, however, according to existing knowledge, that tissue preserved by freezing should be used with the least possible delay after thawing.

TABLE 4. Viability of Ovarian Tissue *in vitro* after Freezing

Medium in which ovaries frozen	Treatment of graft after thawing	No. of implantations	No. of functional grafts obtained	Average time taken for reappearance of cycle (days)
Glycerol-saline	Implanted immediately	20	20	19.2
Glycerol-serum	Implanted immediately	19	19	15.1
Glycerol-saline	Left for 3 hrs. at room temperature	20	2	25.5
Glycerol-serum	Left for 3 hrs. at room temperature	19	13	19.6
Glycerol-saline	Left for 3 hrs. at 37° C.	10	0	0
Glycerol-serum	Left for 3 hrs. at 37° C.	10	1	23.0
Glycerol-saline	Left for 3 hrs. at 2° C.	9	7	20.1
Glycerol-serum	Left for 3 hrs. at 2° C.	10	5	19.4

for prolonged periods, however, will depend largely on its use for homografting, and this fact involves some consideration of the practicability of making grafts from one individual to another of the same species. Much of the medical thought on this matter has been coloured by the well-known fact that it is almost impossible to make homografts of skin, the tissue most commonly grafted, although it is easy to make autografts. By contrast, the experimental biologist interested in the grafting of gonads has encountered little difficulty in securing 'takes' of gonadal tissue in a different individual. The classic work, for instance, of

a kidney, and one or four such halves for subcutaneous grafting. The freezing procedure found most successful for ovarian tissue was generally used, that is to say soaking in 15 per cent glycerol-saline or serum and slow cooling to $-79°$ C. or $-190°$ C. The test procedure was to maintain the animal on the stock diet plus saline to drink for two or three weeks, and then transfer it to a salt-free diet for two weeks. In some experiments an intermediate period on stock diet only, containing a small amount of salt, was introduced between the added salt and the salt-free regimens, making a six weeks' test period in all. If the rat survived the salt-free period in good health, and was found at autopsy to have a visible graft and no accessory or residual adrenal tissue, it was considered that the graft was active. All recovered grafts were sectioned. The results now available for a large number of animals are very mixed.

TABLE 6. Grafting Experiments with normal and frozen Adrenal Cortex (Subcutaneous Grafts of four Enucleated Half-adrenals)

	Total no. of animals	Discarded as incomplete	Dying during experiment	Surviving salt-free period
Untreated cortex	30	1	7	22
Glycerol-treated cortex	30	0	4	26
Frozen glycerol-treated cortex	30	2	21	7

The best results with both normal and treated cortex were obtained when the amount of tissue was increased by the implantation of four enucleated halves instead of one, and are summarized in Table 6. It will be seen that treatment with glycerol by present methods, without freezing, apparently affects very little the capacity of cortical tissue to make a graft, but freezing greatly diminishes the proportion of 'takes'. Histologically, grafts developed from frozen tissue are similar to those derived from normal tissue (Plate XXIV, Fig. 7).

Testis

The information obtained for the rat ovary was found to be readily applicable to the rat testis. The testes were taken from male rats seven to nine days old (at which age the organ is firm and easy to handle) and after various treatments were grafted to castrated adults. The homografts took well and freezing the tissue for up to seven days at $-79°$ C. did not seem to make any difference to its ability to form a functional graft as judged by its morphological condition, the size of the seminal vesicles and the behaviour of the animal. Our data for longer periods of freezing are not yet complete, but we already know that testis tissue frozen for up to 54 days may form an endocrinologically active graft; we cannot yet say that it does so with the same regularity as does tissue frozen for shorter periods. Histologically, the grafts, frozen and unfrozen, were typical of those usually obtained in the subcutaneous position (Plate XXIV, Fig. 6). The tubules were small and occupied a relatively small part of the testis. The interstitial and other inter-tubular tissue was prominent. The later stages of spermatogenesis were absent, but some tubules, even in the grafts of frozen and thawed tissue, contained spermatocytes in active division. We may conclude, therefore, that testis tissue is relatively resistant to low temperatures under the conditions employed by us, and it remains to evolve a technique which will permit us to investigate the later stages of spermatogenesis after freezing.

Adrenal Cortex

Experiments with adrenal cortical tissue have so far been much less promising than those with ovary and testis, partly no doubt because of the much greater complexity of the experiments involved. It is generally agreed that a graft of normal adrenal cortex will often 'take' and maintain life in the adrenalectomized animal, provided the animal is maintained on salt while the graft is becoming established, but there is clearly much scope for variation of grafting technique and test procedure. In all our earlier experiments the adrenals were removed, and cut into halves, which were squeezed to remove the medulla and inner cortex. A single half, so treated, was used for grafting into

increase the interstices between them, but that it certainly does not prevent their formation. The protective effect of glycerol

Apart from crystallization of the water it is not difficult to recognize other possible causes of damage during freezing. Breedis (1942), interested in the freezing of tumour cells, had already called attention to the probability that the concentration of salt was largely responsible for such damage. As the water freezes out of a biological fluid, substances in solution, notably the electrolytes, become more concentrated in the residual fluid, which therefore becomes progressively more hypertonic and damaging to the cell. Thus, during freezing, the cell is in danger from the time water begins to freeze out until the material becomes completely solid or until the temperature gets sufficiently low to render negligible over ordinary periods the physical effects of the hypertonic residual fluid. Recent studies by my colleague Dr. Lovelock have provided precise data in support of this view. In the case of red blood-cells, for instance, the damage caused by freezing under various conditions can be paralleled exactly at normal temperatures by exposing the red cells to concentrations of NaCl corresponding to those produced in the residual fluid by the freezing process (Lovelock, 1953).

This view implies that the quicker the freezing process the shorter the danger period, and offers an alternative explanation of the beneficial effects of ultra-rapid freezing noted by Luyet. The fact that even the instantaneous vitrification of minute amounts of material met with only limited success, especially with spermatozoa (Hoagland and Pincus, 1942), is probably due to the occurrence of thermal shock. It has been known for a long time that bull spermatozoa and other cells are seriously damaged by sudden falls of temperature in the range above zero, and the same thing happens in the range below zero (Lovelock, 1953). The rate of freezing, to be optimal, must therefore be slow enough to avoid thermal shock, and fast enough to avoid adverse effects from the hypertonic residual fluid.

Anterior Pituitary Tissue

Experiments being carried out by Mr. S. E. Smith show that anterior pituitary tissue of the rabbit will grow well in culture, and that the capacity to do so is not usually abolished by freezing and storage at $-79°$ C.

Freezing under the Microscope

The discovery of the remarkable properties of glycerol in protecting living cells subjected to cooling naturally raised the

a special freezing slide was designed (Smith, Polge and Smiles, 1951). The slide consisted essentially of a block of perspex with a chamber in the middle to accommodate the specimen, and fitted with a thermocouple to register the temperature, and two copper side-arms to dip into liquid air for cooling or into warm water for warming. Two photomicrographs obtained with the aid of this apparatus are shown in Plate XXV, Fig. 8. (At this stage the film entitled 'Freezing and Thawing of Living Cells'—Smith, Polge and Smiles—was shown in the lecture.)

Discussion

Nature of the Damage caused by Freezing and of the Protective Action of Glycerol

I have already referred to the idea that freezing damages cells mechanically by the formation of ice crystals, and that if a cell is to survive, cooling and re-warming must be ultra-rapid so that vitrification rather than crystallization of the water takes place. This idea was supported superficially by the fact that Luyet and others had obtained their best results by the very rapid freezing of very small amounts of material. On the other hand, at the time we started work, other well-authenticated observations were difficult to reconcile with the vitrification theory. Since that time our work with glycerol seems to have disposed of it. You have seen that, so far as the medium is concerned, the presence of glyercol may alter the shape of the crystals, and may

aged. It must be supposed, therefore, that the cells which succumb during the period at low temperature do so for some other reason than biochemical ageing (Parkes, 1952a). This reason must presumably depend on some physical effect associated with the complex interaction of cell, medium, and temperature. Such an effect, however, is not inherent in the problem, and may well be avoidable. It appears, for instance, that the bull sperm and ovarian tissue, under the optimal conditions used by us, are in a stable state at −79° C. and −190° C. respectively. The factors involved in securing such a stable state are evidently of prime importance and their elucidation for other cells will be of much interest.

If, as seems likely, it will be possible to eliminate physical effects from the storage problem, interest will shift back to the metabolic side. On the basis of experiments lasting only a year it is impossible to postulate complete arrest of ageing of the cell. If, for the sake of argument, we assume that biochemical activity is halved for each drop of 10° C., then activity at −79° C. would be about 1/4000th that at +40° C. In other words, if ageing depends on biochemical activity, a cell living one day *in vitro* at 40° C. would live more than 10 years at −79° C. At this rate a bull spermatozoon might live 10 to 15 years, and a red blood-cell a great deal longer. Factors other than temperature coefficients would probably extend such periods almost indefinitely, and the experimental investigation of such potential immortality is clearly a long-term project. Nevertheless, in the last year, largely under the influence of Dr. Lovelock, we have

[illegible]

REFERENCES

BLUMENTHAL, H. T. and WALSH, L. B. (1950). *Proc. Soc. exp. Biol., N.Y.* 73, 62

BREEDIS, C. (1942) *J. exp Med.* 76, 221

In tl ' . . .. ' ... mgent
po: ' . . protective action of glycerol.
Fir, ence of the glycerol lowers the temperature at which a particular degree of hypertonicity becomes operative. Secondly, the presence of the glycerol increases the amount of the residual fluid and, since NaCl is soluble in glycerol, will again reduce the salinity. During thawing, the same events happen in the reverse order, and by ordinary methods of ng the inversion of the temperature ng the time for . . make the process of th ng.

If current e possible or avoid keeping the electrolyte concentra- t ow as possible. Quite evidently a great deal of information has still to be obtained about freezing damage and methods of avoiding it, especially in relation to the extraordinary differences between different types of cell pects are beginning to f this ll to much

Survival in Low Temperatures

At the beginning of this lecture I referred to the theoretical expectation that if a cell can be made to survive freezing and thawing, the interval between these processes should be immaterial because biochemical changes will be arrested or reduced to a negligible level at low temperatures. In general our work accords with this view, but at the t the view is very
i long-term storage in be- t the fowl spermatozoa and human red cells in experiments so far reported there is a small but steady loss of cells as the time at low temperature is prolonged, but the cells that survive are in the same functional state as when the period of suspended animation began, i.e. they do not appear to have

XV

The Life-span of Red Blood-cells

P. L. MOLLISON

In almost all text-books of physiology written before 1945 the life-span of the red cell was given as about 30 days; text-books of haematology expressed uncertainty on this subject and gave no estimate of red cell life-span in disease. During the last ten years knowledge in this field has advanced very rapidly. The average life-span of red cells in normal subjects has been determined by several different methods with a relatively small scatter of results, all estimates falling within the range of 109–127 days. Moreover the life-span of red cells in numerous blood diseases has been measured and in some cases this has given a much better insight into the mechanism of the anaemia. This new knowledge has been gained almost entirely by experiments in which red cells are 'tagged' in some way. For reasons which I shall discuss later, the bulk of the knowledge has in fact been gained from observations following transfusions in which the donor's red cells are distinguished by serological methods from those of the recipient.

However, before dismissing past estimates of red cell life-span I should like to discuss the observations on which they were based and to explain how difficult it was to interpret the evidence precisely.

Early Estimates

The estimate that red cells lived for 30 days was based partly upon experiments in which subjects developed plethora as a result of residence at an altitude of 10,000 feet above sea level and then returned to sea level again. It was found that on moving to 10,000 feet the red cell count rose by one to one and a

DAVENPORT, C. B. (1897). *Experimental Morphology:* Part I. Effect of Chemical and Physical Agents upon Protoplasm. Macmillan Co., New York.
HOAGLAND, H. and PINCUS, G. (1942). *J. gen. Physiol.* 25, 337.
LOVELOCK, J. E. (1952). *Lancet*, June 21, 1238.
LOVELOCK, J. E. (1953). *Biochim. biophys. Acta*, in press.
LUYET, B. J. and HODAPP, E. L. (1938). *Proc. Soc. exp. Biol., N.Y.* 39, 433.
MEDAWAR, P. B. and BILLINGHAM, R. E. (1952). *J. exp. Biol.* 29, 454.
MOLLISON, P. L. and SLOVITER, H. A. (1951). *Lancet*, November 10, 862.
MOLLISON, P. L., SLOVITER, H. A. and CHAPLIN, H. Jnr. (1952). *Lancet*, September 13, 501.
PARKES, A. S. (1945). *Brit. med. J.* ii, 212.
PARKES, A. S. (1952a). *Freezing and Drying*, p. 99. Publ. by Institute of Biology.
PARKES, A. S. (1952b). *Ciba Foundation Colloquium on Endocrinology*, IV, 264.
PARKES, A. S. and SMITH, A. U. (1953). *Proc. Roy. Soc.*, B, 140, 455.
POLGE, C. (1951). *Nature, Lond.* 167, 949.
POLGE, C. and PARKES, A. S. (1952). *Anim. Breed. Abstr.* 20, 1.
POLGE, C. and ROWSON, L. E. A. (1952a). *Nature, Lond.* 169, 626.
POLGE, C. and ROWSON, L. E. A. (1952b). *Vet. Rec.* 64, 851.
POLGE, C., SMITH, A. U. and PARKES, A. S (1949). *Nature, Lond.* 164, 666.
SHAFFNER, C. S. (1942). *Science*, 96, 337.
SHAFFNER, C. S., HENDERSON, E. W. and CARD, C. G. (1941). *Poultry Sci.* 20, 259.
SLOVITER, H. A. (1951) *Lancet*, April 14, 823.
SLOVITER, H. A. (1952). *Nature, Lond.* 169, 1013.
SMITH, A. U (1949). *Nature, Lond.* 164, 1136.
SMITH, A. U (1950). *Lancet*, December 30, 910.
SMITH, A U (1952a). *Nature, Lond.* 170, 374.
SMITH, A U. (1952b). *Exp Cell Research*, 3, 574.
SMITH, A. U. and PARKES, A. S. (1951). *Lancet*, September 29, 570.
SMITH, A U. and POLGE, C. (1950a). *Nature, Lond.* 166, 668.
SMITH, A. U. and POLGE, C (1950b) *Vet. Rec.* 62, 115.
SMITH, A. U., POLGE, C. and SMILES, J (1951). *J Roy. Micr. Soc.* 71, 186.

Thus this method of estimating life-span by producing plethora and then noting the time for which it persists will in fact give a good estimate of the life-span of the red cell only if red cell production remains constant throughout the period of the observations. In some patients in whom red cell production is already diminished, transfusion may fail to depress production any further and thus after transfusion packed cell volume may be increased for a period as long as three months. For example the observations shown in the lower half of Fig. 1 were made on a patient with chronic nephritis. After transfusion packed cell volume was raised above its pre-transfusion level for as long as 94 days. Such observations suggest that the life-span of red cells cannot be less than 94 days although this assumes that red cell production does not increase during the period of the observations.

The other observations which were wrongly interpreted in the past were those of bile pigment excretion. The method was to establish bile fistulae in animals and thus to collect all the bile pigment excreted. However, such animals usually fell into ill health and variable results were obtained. For example, the excretion of bile pigment might be temporarily reduced due to liver dysfunction following operation. Later there might be an outpouring of this stored-up pigment. If measurements of excretion were now made, a gross overestimate of red cell turnover would result. Hawkins and Whipple (1938) managed to improve the technique and keep their animals in good health. Moreover, instead of simple measurement of the output of bile pigment they adopted a more ingenious plan. They bled the dog or gave it phenylhydrazine and thus produced a sudden anaemia. Shortly afterwards the circulation was flooded with young red cells; about four months later there was an increase in the excretion of bile pigment. Assuming that the average red cell was born 15 days after the blood loss and died at the peak of the bilirubin excretion it could be concluded that the life-span of dog red cells was 124 days. The same workers found that when 1 gm. of haemoglobin was injected intravenously an additional 40 mgm. of bile pigment was recovered from the fistula. Using this figure and measuring the daily output in normal dogs they again

half million red cells per c.mm., and on return to sea level the count fell again about in 20 days to its normal level (Escobar and Baldwin, 1934). The same sort of changes are observed when normal subjects are rendered plethoric by transfusion. For example Pace, Lozner *et al.* (1947) gave transfusions of 1,000 cc. of red cells to a series of normal subjects and raised

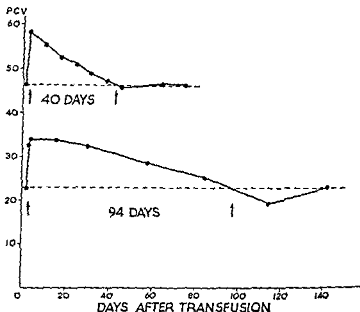


FIG. 1. Changes in packed cell volume (P.C.V.) following transfusion. Upper curve: a group of normal subjects (data from Pace *et al.*, 1947). Lower curve: a single subject with chronic nephritis. In both curves the arrows mark the departure from and return to a base line.

the packed cell volume on the average from 46.5 to 58.5 per cent. The packed cell volume remained above its pre-transfusion level for about 40 days as is shown in Fig. 1. Now if the production of red cells remained constant, the packed cell volume would remain above its initial level for as long as any transfused red cells were left. However, in plethora, red cell production diminishes as shown by a reduction in the reticulocyte count (Pace *et al.*, 1947, Robertson, 1917); therefore the only safe deduction is that the added red cells live for more than 40 days.

thought up to that time that tests made *in vitro*, in particular measurements of the amount of spontaneous haemolysis and of osmotic fragility, indicated the state of preservation of the red cells. However, transfusion experiments showed that such criteria were unreliable. For example, red cells stored in citrate-sucrose remain virtually unhaemolysed after two weeks' storage and their osmotic fragility is actually lower than normal; however if these red cells are used for transfusion they are very rapidly removed from the recipient's circulation (Mollison and Young, 1941, 1942). Maizels (1941, 1943) was one of the first to point out that the test of osmotic fragility is particularly misleading if used as a guide to the state of viability of stored red cells since it reflects the composition of the preservative fluid rather than the state of preservation of the cells.

Up to the present time measurement of the survival of transfused red cells continues to be the main criterion of viability of stored cells. For example, Halpern *et al.* (1950) found that a phenothiazine derivative, closely related to phenergan, would delay the rate of spontaneous lysis and of potassium loss from stored cells and they tentatively concluded that it was a valuable red cell preservative. However, measurements of survival *in vivo* showed that this substance does not increase the viability of stored cells in the least (Chaplin *et al.*, 1952).

The successful application of the method of differential agglutination to the problem of red cell preservation reawakened interest in the method. It was soon realized that accurate results could be obtained only if really complete agglutination of the recipient cells could be achieved and one aid to this was the use of centrifugation (Dacie and Mollison, 1943). However the most important factor of all is the preparation of extremely avid agglutinating sera, and there can be no doubt that sera produced in animals are far more suitable than any reagents that can be obtained from humans. With sera produced in animals it is possible to agglutinate all but a few thousand per c.mm. of the recipient's own red cells. For the production of anti-A, the method of Morgan (1943) is the best. The last ingredient in the recipe for success is the patience to count large numbers of red cells.

concluded that the life-span was about 120 days. Subsequent estimates have substantially confirmed this figure. A variation of this method was employed by Harne and others (1945). Rhesus monkeys were bled and frequent reticulocyte counts were made. The appearance of showers of reticulocytes 94 to 107 days after haemorrhage indicated that the cells formed after the haemorrhage were dying out and being replaced by new ones.

However, all these methods mentioned so far are rather indirect and methods which involve tagging red cells in one way or another are much more satisfactory. As already mentioned, much of our knowledge about the life-span of red cells in man comes from transfusion experiments and it is with such experiments that I wish to deal most extensively.

Transfusion

The first workers to attempt to use transfusion to measure red cell survival were Todd and White (1911). They prepared sera which would haemolyse the red cells of one animal but not of another. Thus after transfusion they could take samples from the recipient animal and analyse them by, for example, haemolysing the red cells of the recipient and thus leaving intact the red cells of the donor for counting. Todd and White could draw no conclusions on the normal life-span of red cells in cattle because these transfused cells acted as antigens; immune antibodies were rapidly produced and the transfused red cells survived for only a few days.

This principle of differentiating between the cells of the donor and recipient was applied to man by Ashby (1919) who took advantage of the fact that transfusions of group O blood were quite often given to recipients of group A or B. Ashby showed that if samples were taken from a group A patient who had received some group O blood and red cell counts were made using anti-A serum as a diluent almost all the patient's own cells could be agglutinated but those of the donor left free for counting. Very little use was made of this method until the early 1940s when it came to be realized that measurement of the survival of transfused red cells *in vivo* was the only reliable criterion of the state of preservation of red cells. For example it was

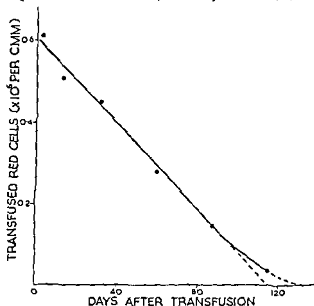
Labelled Glycine

Transfusion experiments have often been severely criticized on the grounds that they deal necessarily with red cells of all ages and that conclusions from them are based on inference; for example, it is assumed that the transfused red cells survive in just the same way as the recipient's own cells. Nevertheless, conclusions drawn from transfusion experiments have received confirmation from many sources in the last few years. Two main methods have been used in an attempt to produce a tagged population of cells of all about the same age. Shemin and Rittenberg (1946) fed glycine labelled with N^{15} to a normal subject and noted that the labelled glycine was incorporated in newly formed haemoglobin. After about 30 days the concentration of labelled glycine reached its maximum and remained at this level for approximately 70 days. The concentration then fell rapidly. The time between the mid-point of the up-curve and the mid-point of the down-curve was 127 days and this was taken to be the average life-span of the red cells. Similar observations were reported by Gray and Neuberger (1950).

Although these observations provide valuable support for evidence obtained by other methods, some points in their interpretation may be criticized. First of all, the uptake time is extremely long, labelled cells being produced for three weeks or more, so that the tagged cells are not in fact all of the same age. Because of this, labelled cells disappear over an extended period and this interferes with the precise determination of the mean life-span and in the estimation of the scatter of life-span. Another serious disadvantage of this method is that there appears to be some re-utilization of labelled glycine. In some of Shemin and Rittenberg's (1946) curves the amount of labelled haem present in the blood 180 days after transfusion amounted to as much as one-third of the maximum level reached soon after feeding the labelled glycine. There also remains some doubt whether labelled glycine can leave intact red cells. That it may possibly do so is suggested by the observation that one-quarter of the labelled cells 'disappear' less than 106 days after labelling (London, Shemin, West and Rittenberg, 1949). Since other

NORMAL SUBJECTS

When normal blood is transfused to relatively normal subjects the estimates of the number of surviving cells when plotted against time fall near to a straight line which cuts the time axis at about 120 days (Fig. 2). From this it is deduced that the average life-span of the red cells, that is, the time between their



release from the bone marrow of the donor and their final disappearance from the blood stream of the recipient, is 120 days. Curves of the survival of transfused red cells often show a little tail at the end—a suggestion of this may be seen in Fig. 2. Dornhorst has pointed out (1951) that such a tail is expected if red cells do not all live for exactly the same time. In fact if the shape of the tail could be accurately defined the standard deviation of life-span could be deduced From such data as exist the standard deviation has been calculated to be approximately 6 days (Mollison, 1951).

relatively long half-life the radiation hazard must be carefully considered. For these various reasons very few estimates of red cell life-span in man have been made by this method.

Thus neither of these isotope methods lends itself well to clinical use, and this explains why we must still rely heavily on the method of differential agglutination for our knowledge of red cell life-span in man. During the last ten years the use of this method has contributed considerably to our understanding of the nature of anaemias and its possibilities do not yet seem to be exhausted.

Differential Agglutination

It was in hereditary spherocytosis that the first really striking results were obtained with the method of differential agglutination. It was found that normal red cells transfused to patients with this disease survived normally (Dacie and Mollison, 1943). Up to this time there was a considerable body of opinion which held that there was an abnormal mechanism in these patients which destroyed red cells, and these experiments showed this view to be false.

As expected, when red cells were taken from patients with hereditary spherocytosis and transfused to normal subjects, they survived for a much shorter time than normal (Dacie and Mollison, 1943; Loutit and Mollison, 1946). Under these circumstances it is reasonable to conclude that the fault lies in the red cells and that the change of environment brought about by the transfusion does not affect their survival. In that case the initial slope of the survival curve, extrapolated, cuts the base line at the mean cell life (Dornhorst, 1951). For this to be a practical means of estimation, the initial slope must be well defined and this is sometimes the case (see Fig 19 in Mollison, 1951).

In special circumstances the transfusion of faulty red cells may alter their survival and this possibility should always be considered in interpreting transfusion experiments. For example, Dacie and Mollison (1949) obtained red cells from a donor with paroxysmal nocturnal haemoglobinuria and transfused them to two recipients, an adult and an infant. The red cells survived far better in the circulation of the infant and tests *in vitro* suggested that this was because the infant's plasma was relatively deficient

evidence suggests that the variation in life-span is comparatively small this disappearance may be of the label rather than the red cell. On the whole this method seems more useful in studying Hb metabolism than in giving precise estimation of red cell life-span.

Radio-iron

A second method of tagging the subject's own red cells is that of injecting radio-iron. As soon as the uptake curve of radio-iron reaches its plateau large intravenous injections of non-radio-active iron are given and further doses are given during the course of the experiment. If this is not done most of the radio-active iron released from breaking-down red cells will be incorporated in newly formed red cells and will interfere with the survival curve (Finch *et al.*, 1949).

This method has the great advantage of tagging the subject's own red cells and in particular of tagging cells all formed during a short period of time. Published curves based on experiments in dogs show a comparatively rapid uptake of radio-iron, complete in about 10 days; there is then no change in circulating radio-activity for about 90 days suggesting a negligible incidence of cell deaths during this period (Brown and Eadie, 1953). From such experiments the standard deviation of life-span was calculated to be 6 days, a conclusion which agrees closely with the rough estimates, mentioned above, made from transfusion experiments in man.

Unfortunately, this method of estimating red cell life-span with radio-active iron can be applied only in a subject in whom re-utilization of iron is minimal and thus satisfactory observations can seldom be made in man. However, subjects with aplastic anaemia or others with grossly impaired red cell production are suitable, and subjects with haemochromatosis can also be used since in them iron stores are enormously increased and re-utilization of radio-iron from disintegrating red cells is minimal (Finch *et al.*, 1949). In all other subjects enormous doses of non-radio-active iron would have to be given following the administration of radio-iron to prevent re-utilization. A further disadvantage of using radio-active iron is that it is excreted from the body in only minimal quantities and as it has a

In pernicious anaemia red cell life-span is only moderately reduced, the published curves of Loutit (1946) and Singer *et al.* (1948) suggesting a mean cell life of about 20 or 30 days. The severe anaemia thus indicates that output is hardly greater than in the non-anaemic subject.

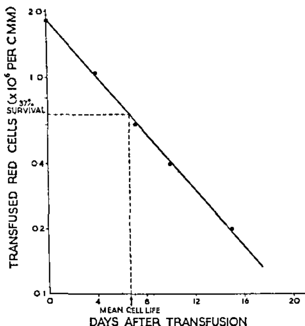


Fig. 4. Survival of transfused red cells, plotted on a log scale, in a patient with moderately severe acquired haemolytic anaemia (Case 2) from Fig. 3 (from Mollison, 1951).

In acquired haemolytic anaemia the patient has a mechanism which destroys all red cells and curves like the ones shown in Fig. 3 are obtained when normal red cells are transfused.

As Dornhorst (1951) has pointed out, provided that almost all the red cells live for less than 20 or 30 days, death by ageing can be ignored. When the results are plotted on a logarithmic scale, linear slopes are often obtained and mean cell life is then given by the time taken to eliminate 63 per cent of the transfused cells. Thus, in the example given in Fig. 4, mean cell life is estimated to

in the lytic factor. The observations in the infant suggest a mean cell life of 13 days whereas those in the adult lead to an estimate of 6 days. The latter estimate fits better with the observation that the donor had a reticulocyte count of 600,000 per c.mm. and a total red cell count of 1·5 million per c.mm.

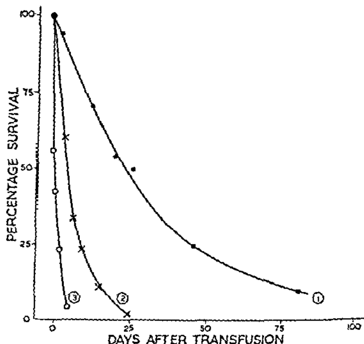


Fig. 3 Survival of transfused red cells in three patients with acquired haemolytic anaemia. (1) mild, (2) moderate; (3) severe (from Mollison, 1951).

Thus if the life-span of the red cell is reduced from 120 days to 6 days, the output of new red cells will have to increase twentyfold if a normal red cell count is to be maintained. To maintain a count of 1 5 million will require only 30 per cent [illegible], to say, six times normal The [illegible] is about 60 [illegible] appeared [illegible] output. Often, in making calculatio[illegible] found that the number of reticulocytes found is r[illegible] greater than expected. This suggests that when the output from the marrow increases reticulocytes may be released at a slightly earlier stage of development and thus spend longer as reticulocytes in the peripheral blood.

output this patient could have maintained a haemoglobin concentration of only about 5 g./100 ml. As her haemoglobin concentration was in fact about 7 or 8 g./100 ml. it can be taken that she was increasing her output a little in response to the anaemia but not nearly as much as a person *with a normal marrow*. From

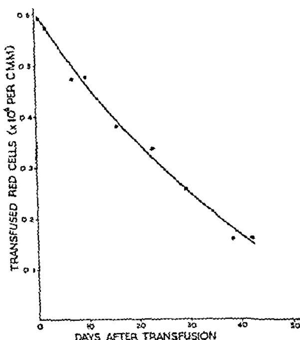


FIG. 5. Survival of transfused red cells in a patient with nephritis in an active phase of the disease (from Mollison, 1951)

these observations alone it is thus safe to conclude that the response *of the marrow to the stimulus of anaemia is grossly* impaired in such patients. In more chronic cases of nephritis one may find a strictly normal survival of transfused red cells and in such cases the diminution in red cell production is presumably indicated precisely by the diminution in packed cell volume. Thus, in the patient from whom the data in the lower half of Fig. 1 were obtained, the survival of transfused red cells was normal, and the patient maintained his packed cell volume constant at

be 7 days. Now if a normal haemoglobin level were to be maintained, an output of about sixteenfold would be required. To maintain a level of about 6 g./100 ml., as this patient did, required an output of only six or seven times normal.

If one examines a series of cases of acquired haemolytic anaemia it is found that in patients with a severe haemolytic process, maintaining haemoglobin concentrations of 5-7 g/100 ml. and red cell counts of 1·5 to 2·0 million per c.mm., mean cell life is of the order of 3 to 5 days and it can be calculated that this demands an output of about eight times normal. This may provisionally be regarded as the upper limit of increased red cell production (cf. Crosby and Akeroyd, 1952).

Some idea of the furious activity of the marrow can be obtained by considering that the normal man discharges about three million new red cells per second into the blood stream and that patients with severe haemolytic anaemia may approach thirty million per second. Expressing these amounts in terms of packed red cells is less dramatic, the normal man forming less than 1 ml. of packed cells per hour and the patient with severe haemolytic anaemia almost 10 ml. per hour.

A group of anaemias which need further study are those in which mean cell life is only slightly diminished, for example, some cases of rheumatoid arthritis, rheumatic fever in relapse, disseminated lupus erythematosus and nephritis in an active stage. Fig. 5 shows the survival of transfused red cells in a case of nephritis. In this type of case, in which an appreciable proportion of transfused red cells live for more than 30 days, the effect of the loss of red cells by ageing cannot be ignored. As Dornhorst (1951) has pointed out, there will be fewer old cells in the patient's circulation than in the donor's for in the patient's circulation cells will not have a normal chance of living for their normal life-span. The effect of this is that the rate of disappearance of transfused red cells from the patient's circulation must be a little greater than the rate of disappearance of the patient's own red cells. A method of allowing for this and deducing the mean life of the patient's own red cells has been described (Dornhorst, 1951). In the particular case shown in Fig. 5 mean cell life is calculated to be approximately 35 days. With normal

while the other lasted for the normal time. This seems to make it almost certain that the phenomenon is after all due to incompatibility, though it has not so far been possible to demonstrate, *in vitro*, any interaction between the red cells of the donor and the serum of the recipient even after the red cells have been

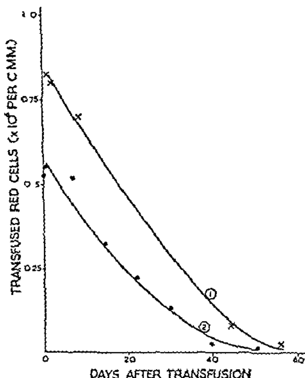


Fig. 6. Incompatibility without demonstrable antibodies—see text (from Mollison, 1951).

eliminated. One reason for believing that incompatibility is the cause is the fact that Dr. L. E. Young, of Rochester, U.S.A., and his co-workers have observed a similar phenomenon in the course of transfusion experiments in dogs (personal communication). Thus in one dog compatible cells at first survived normally but after a subsequent transfusion the rate of elimination was increased, no incompatibility between the donor's cells and

about 22 per cent and may thus be taken to have had 50 per cent normal production.

Although transfusion experiments can give such valuable information about red cell life-span they must always be cautiously interpreted. One example has already been given in discussing the survival of red cells from a patient with paroxysmal nocturnal haemoglobinuria. Red cells from this patient survived far better in one recipient than in another. It is known that red cells from patients with this disorder are haemolysed by the action of components present in normal plasma and the observed difference in survival *in vivo* in the two recipients was paralleled by a difference in lysis *in vitro* when the cells were incubated with serum from the two recipients (Dacie and Mollison, 1949).

Occasionally, blood group incompatibility interferes with survival experiments. If Rh positive cells are transfused to an Rh negative subject, they may survive well initially but may later stimulate the formation of Rh antibody which leads to the rapid destruction of the surviving cells. However such gross forms of incompatibility can be prevented.

A phenomenon which may represent a different kind of incompatibility is the increased rate of elimination of apparently compatible red cells in subjects whose own cells appear to have a normal life-span. This phenomenon was first described by Loutit, Mollison and Young (1943). Curve 1 in Fig. 6 indicates the survival of red cells stored for 14 days in acid-citrate-dextrose, in a woman with an anaemia following an abortion. When these observations were first made it was thought possible that the diminished survival might be an effect of storage. Accordingly fresh blood was collected from the same donor and transfused to the same recipient on a subsequent occasion. As curve 2 shows, survival was almost identical. At the time it was considered that this showed that the phenomenon was not due to incompatibility.

Subsequently at the Blood Transfusion Research Unit a similar phenomenon has quite often been encountered. In one patient in whom red cells from two donors were being followed simultaneously, one population of red cells disappeared rapidly

while the other lasted for the normal time. This seems to make it almost certain that the phenomenon is after all due to incompatibility, though it has not so far been possible to demonstrate, *in vitro*, any interaction between the red cells of the donor and the serum of the recipient even after the red cells have been

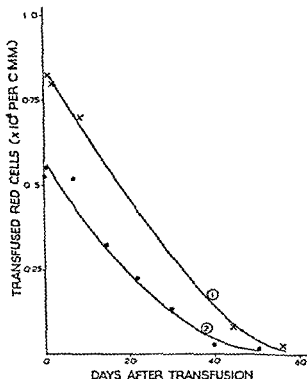


FIG. 6. Incompatibility without demonstrable antibodies—see text (from Mollison, 1951).

eliminated. One reason for believing that incompatibility is the cause is the fact that Dr. L. E. Young, of Rochester, U.S.A., and his co-workers have observed a similar phenomenon in the course of transfusion experiments in dogs (personal communication). Thus in one dog compatible cells at first survived normally but after a subsequent transfusion the rate of elimination was increased; no incompatibility between the donor's cells and

about 22 per cent and may thus be taken to have had 50 per cent normal production.

Although transfusion experiments can give such valuable information about red cell life-span they must always be cautiously interpreted. One example has already been given in discussing the survival of red cells from a patient with paroxysmal nocturnal haemoglobinuria. Red cells from this patient survived far better in *one* recipient than in another. It is known that red cells from patients with this disord... e haemolysed by the action of ... l plasma and th... ...nd the recipient ...ence in lysis *in vitro* when the cells were ...ubated with serum from the two recipients (Dacie and Mollison, 1949).

Occasionally, blood group incompatibility interferes with survival experiments. If Rh positive cells are transfused to an Rh negative subject th... ...itially but may late... ...y which leads to the ... surviving cells. However such gross f...s of incompatibility can be prevented.

A phenomenon which may represent a different kind of incompatibility is the increased rate of elimination of apparently compatible red cells in subjects whose own cells appear to have a normal life-span. This phenomenon was first described by Loutit, Mollison and Young (1943). Curve 1 in Fig. 6 indicates the survival of red cells stored for 14 days in acid-citrate-dextrose, in a woman with an anaemia following an abortion. When these observations were first made it was thought possible that the diminished survival might be an effect of ... - ingly fresh blood was c... fir... ... ...ion. At the time it was con- s... ...s showed that the phenomenon was not due to incompatibility.

Subsequently at the Blood Transfusi... ... simi- lar phenomen... ... one ... ...g followed ... ... of red cells disappeared rapidly

effect. It is desirable that this conclusion should be confirmed by other workers.

Conclusion

It is now recognized that the measurement of red cell life-span in man is a valuable method of investigation. Much more could be done at the present time to make accurate measurement in haemolytic anaemias—particularly in patients maintaining a stable haemoglobin concentration. In such patients it should usually be possible to measure the average life-span of the red cells quite accurately. Such studies would be even more rewarding if there were also a method of measuring the output of new red cells directly. Reticulocyte counts are too indirect for this purpose though possibly newer radio-iron techniques may be suitable. When such a method is developed it may be possible to measure red cell destruction and production accurately in the same patient and thus gain a better understanding of the anaemias.

REFERENCES

Ashby, W. (1919). *J. exper. Med* 29, 267.

Brown, I. W. and Eadie, G S (1953). *J. gen. Physiol.* 36, 327.

Callender, S. T, Powell, E O. and Witts, L. J. (1947). *J Path. Bact.* 59, 519.

Chaplin, H., Crawford, Hal, Cutbush, Marie and Mollison, P. L. (1952) *J. clin Path* 5, 91.

Crosby, W H. and Akeroyd, J. H (1952) *Amer. J Med.* 13, 273

Dacie, J. V. and Mollison, P. L. (1943) *Lancet*, 1, 550.

Dacie, J. V. and Mollison, P. L. (1949) *Lancet*, 1, 390.

Dornhorst, A. C. (1951). *Blood*, 6, 1284

Escobar, R A and Baldwin, F M (1934) *Amer. J. Physiol.* 107, 249.

Finch, C. A, Wolff, J A., Rath, C. E and Fluharty, R G (1949). *J. lab. clin Med.* 34, 1480.

Gray, C H. and Neuberger, A (1950). *Biochem J.* 47, 81.

Halpern, B. N, Dreyfus, B. and Bourbon, G. (1950). *Pr. med* 58, 1151.

Hahn, O G, Lutz, J F., Zimmerman, G. I. and Davis, G L. (1945). *J. lab clin Med* 30, 247

Hawkins, W. B. and Whipple, G H. (1938). *Amer. J. Physiol.* 122, 418

London, I M., Shemin, D., West, R. and Rittenberg, D (1949) *J biol. Chem* 179, 463

the recipient's serum could be demonstrated. Incidentally in other experiments the same workers found that red cells which were demonstrably incompatible were not necessarily rapidly eliminated. For example, a dog immunized to the dog antigen A was transfused with red cells containing the 'low-grade' antigen A'—roughly comparable with the D^u antigen of the Rh system. The red cells were only slowly eliminated and the curve of elimination resembled the curves shown in Fig. 6.

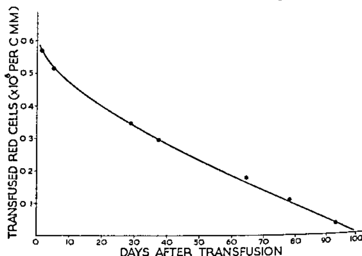


Fig. 7. Survival of transfused red cells in an adult woman (from Mollison, 1951).

These observations suggest that our knowledge of red cell incompatibility in man is not yet complete. It should be added that this phenomenon of incompatibility without demonstrable antibodies is not encountered more frequently than about once in every twenty transfusions.

There is one other unsolved problem in transfusion experiments; that is the diminished survival found in women. Fig. 7 is a typical example. Curvilinear slopes seem to be found only in women who are menstruating; the obvious explanation is that the curvilinearity is due entirely to the loss of red cells in the menstrual discharge. However, Callender, Powell and Witts (1947) who first drew attention to this phenomenon calculated that the known loss of menstruation would not have so great an

effect. It is desirable that this conclusion should be confirmed by other workers.

Conclusion

It is now recognized that the measurement of red cell life-span in man is a valuable method of investigation. Much more could be done at the present time to make accurate measurement in haemolytic anaemias—particularly in patients maintaining a stable haemoglobin concentr[illegible] [illegible]nts it should usually be [illegible] [illegible]an of the red [illegible] [illegible]ould be even more reward-i[illegible] [illegible]s also a method of measuring the output of new red cells directly. Reticulocyte counts are too indirect for this purpose though possibly newer radio-iron techniques may be suitable. When such a method is developed it may be possible to measure red cell destruction and production accurately in the same patient and thus gain a better understanding of the anaemias.

REFERENCES

ASHBY, W. (1919) *J exper Med.* **29**, 267.
BROWN, I. W. and EADIE, G. S. (1953). *J gen. Physiol* **36**, 327
CALLENDER, S. T, POWELL, E. O. and WITTS, L. J. (1947) *J. Path Bact.* **59**, 519.
CHAPLIN, H, CRAWFORD, HAL, CUTBUSH, MARIE and MOLLISON, P. L. (1952). *J clin Path* **5**, 91.
CROSBY, W. H and AKEROYD, J H (1952). *Amer. J. Med.* **13**, 273.
DACIE, J. V. and MOLLISON, P. L. (1943). *Lancet*, **1**, 550
DACIE, J. V. and MOLLISON, P L (1949) *Lancet*, **1**, 390
DORNHORST, A C (1951). *Blood*, **6**, 1284
ESCOBAR, R. A and BALDWIN, F. M. (1934). *Amer J Physiol* **107**, 249.
FINCH, C A, WOLFF, J A., RATH, C. E. and FLUHARTY, R G. (1949). *J. lab. clin Med.* **34**, 1480.
GRAY, C. H and NEUBERGER, A. (1950). *Biochem J* **47**, 81.
HALPERN, B. N, DREYFUS, B and BOURBON, G (1950) *Pr med* **58**, 1151
HARNE, O. G, LUTZ, J. F., ZIMMERMAN, G. I. and DAVIS, C L. (1945). *J. lab. clin. Med* **30**, 247
HAWKINS, W B and WHIPPLE, G. H (1938). *Amer. J. Physiol* **122**, 418.
LONDON, I M, SHEMIN, D., WEST, R and RITTENBERG, D (1949) *J. biol. Chem.* **179**, 463

Loutit, J. F. (1946). *Proc. Roy. Soc. Med.* **39**, 757.
Loutit, J. F. and Mollison, P. L. (1946). *J. Path. Bact.* **58**, 711.
Loutit, J. F., Mollison, P. L. and Young, I. M. (1943). *Quart. J. exper. Physiol.* **32**, 183.
Maizels, M. (1941). Unpublished Report to the Medical Research Council.
Maizels, M. (1943). *Quart. J. exper. Physiol.* **32**, 143.
Mollison, P. L. (1951). *Blood Transfusion in Clinical Medicine.* Blackwell Scientific Publications, Oxford.
Mollison, P. L. and Young, I. M (1941). *Brit. med. J.* ii, 797.
Mollison, P. L. and Young, I. M. (1942). *Quart. J. exper. Physiol* **31**, 359.
Morgan, W. T. J. (1943). *Brit. J. exper. Path.* **24**, 41.
Pace, N., Lozner, E. L., Consolazio, W. V., Pitts, G. C. and Pecora, L. J.
[illegible]
[illegible] **166**, 627.
Singer, K., King, J. C. and Robin, S. (1948). *J. lab clin. Med.* **33**, 1068.
Todd, C. and White, R. G. (1911). *Proc. Roy. Soc.* B. **84**, 255.

XVI

Human Haemoglobins

J. C. WHITE

THE haem-proteins are universally concerned in the respiration of living cells. By virtue of the reversible union with molecular oxygen, haemoglobin is responsible for the respiratory function of the blood, and this has attracted much physiological and clinical research. From the standpoint of protein structure, the haemoglobin molecule is particularly well suited for study.

Recently it has become apparent that fine variations in the structure of haemoglobin may play a part in various human anaemias, and the concept of such conditions as sickle-cell anaemia as a 'molecular disease' has been advanced (Pauling, Itano, Singer and Wells, 1949). This lecture is concerned with the evidence that various haemoglobins may exist in man, and particular attention will be paid to the physico-chemical basis of methods for their demonstration.

THE SPECIFICITY OF HAEMOGLOBIN

Although the haem is the same for haemoglobins from all sources, the basic globin protein exhibits small but significant differences: (*a*) between species, (*b*) between foetal and adult stages in some single species, (*c*) sometimes more than one haemoglobin may co-exist in the post-natal erythrocytes of a single species, or a variant of the 'normal' haemoglobin occur alone. Barcroft devotes a chapter of his monograph on 'Haemoglobin' (1928) to this subject, and draws particular attention to three aspects

(*a*) The extensive crystallographic studies of Reichert and Brown (1909) on vertebrate haemoglobins pointed to great

variations between species as to crystal habit and class, solubility and ease of crystallization (Landsteiner and Heidelberger, 1923). Further, the blood of a single animal (e.g. baboon) might yield several crystallographically distinct haemoglobins. On the other hand, among closely related species, the same sets of crystal systems and groups tend to appear (isomorphism).

(*b*) The positions of the α and β maxima of the visible absorption spectrum are species specific. The span in Ångstrom units which separates the α bands of HbCO and HbO_2 is similarly specific. *Using the Hartridge reversion spectroscope, Munro* Fox (1945) has also found significant variations in the position of the α band in individual rabbits, though not in earthworms and frogs, or in human haemoglobins of various races.

(*c*) The Specific Oxygen Capacity of haemoglobin varies widely in different species (Douglas, Haldane and Haldane, 1912; Maçela and Seliškar, 1925).

In our discussion on the characterization of human haemoglobins, particular attention will be paid to the differences between adult and foetal forms, and to the physical identification of such variants of adult haemoglobin as the sickle-cell haemoglobins.

Foetal and Adult Human Haemoglobin

[illegible]

may be given conveniently in a list modified from the review of Lecks and Wolman (1950), and then considered step by step:

(*a*) Crystallographic differences
(*b*) Solubility differences
(*c*) Differences in amino-acid composition
(*d*) Immunological specificity
(*e*) Greater stability of foetal haemoglobin towards alkali denaturation
(*f*) Rate of spreading in surface films
(*g*) *Electrophoretic mobilities*
(*h*) Spectrophotometric differences in the ultra-violet
(*i*) Differences in the Oxygen Dissociation Curves
(*j*) Clinical significance.

CRYSTALLOGRAPHY OF ADULT HUMAN HAEMOGLOBIN

It has long been known that human haemoglobin could be crystallized (Hunefield, 1840; Funke, 1852), and a number of studies have appeared (Reichert and Brown, 1909; Amantea, 1924; Perrier and Janelli, 1931; Haurowitz, 1935). Several recent studies, however, include also the detailed solubility data necessary for the characterization of homogeneous proteins (Green, Cohn and Blanchard, 1935; Drabkin, 1945, 1946; Jope and O'Brien, 1949; and further refs. below).

Jope and O'Brien (1949) have crystallized adult and foetal haemoglobins as HbO_2, HbCO and methaemoglobin, and as the reduced haemoglobin. Dialysis against ammonium sulphate or potassium phosphate solutions was carried out in the cold in a suitable atmosphere. Adult HbO_2, HbCO and methaemoglobin crystals were isomorphous, belonging to the orthorhombic system, whereas the reduced adult haemoglobin was monoclinic. By a slightly different technique, adding further salt to a clear HbO_2 solution concentrated by dialysis Drabkin (1945, 1946) had prepared bipyramids of human oxy-haemoglobin, which were recognized as belonging to the tetragonal system by optical crystallographic analysis with the polarizing microscope.

The method of X-ray crystallographic analysis of proteins (Kendrew and Perutz, 1949) by the school of Bragg and Perutz at the Cavendish Laboratory, Cambridge, is extending knowledge of the crystal forms of haemoglobin to an amazing extent, and already a 'low-power' view of the external form of the molecule has emerged (Bragg and Perutz, 1952, a b and c; Bragg, 1952). The dry molecule is somewhat egg-shaped, about 47 Å in diameter and 57 Å long. Hydrated, the protein molecule is associated with a molecular layer of bound water, in addition to the surrounding liquid of crystallization in the unit cell. The amount of the latter may vary with salt concentration of the medium in which the crystal is suspended (Perutz, 1946). The bound water increases the volume of the hydrated molecule to more than 100,000 $Å^3$ (Bragg and Perutz, 1952c).

A wide range of mammalian haemoglobins has been analysed by X-ray crystallography, and the results indicate packing in

the unit cells of hydrated molecules of dimensions approximately 55 × 55 × 65 Å. Confirmation has been obtained that adult human HbO_2, and HbCO and methaemoglobin can crystallize isomorphously in either orthorhombic or tetragonal systems. One orthorhombic crystal is of Space group $P2_12_12_1$ (Perutz and Weisz, 1947; Bragg and Perutz, 1952b), with 4 molecules per wet unit cell of volume 564,000 Å³. A second form with slightly smaller dry unit cell (338,000 Å³ instead of 393,000 Å³) is probably of the same Space group, but differs in crystal habit ({001} instead of {110} and {111}), and in being strongly pleochroic (Bragg and Perutz, 1952b). These orthorhombic crystals are polymorphic modifications of the internally symmetrical tetragonal crystal form described by Drabkin (1945, 1946). The tetragonal crystal is a bipyramid, pleochroic and of form {102}; the Space group is $P4_12_1$, the wet unit cell containing 4 molecules in volume 552,000 Å³ (Perutz, Liquori and Eirich, 1951; Bragg and Perutz, 1952b).

The X-ray crystallography of adult human reduced haemoglobin has been described by Liquori, Trotter and Perutz (see Bragg and Perutz, 1952b) for two monoclinic forms, dominant habits {001} and {011} respectively, of Space group $P2_1$; both are strongly pleochroic. The {001} form contains 4 molecules in the wet unit cell of volume 700,000 Å³, whilst the second form has 2 molecules per wet unit cell of volume 272,000 Å³. This form also provided evidence for a close similarity between the structure of human and horse haemoglobin molecules (Bragg and Perutz, 1952b).

Pleochroism has been mentioned several times. This is an optical property of substances arranged in orderly molecular array, and strongly absorbing light in some part of the spectrum. When viewed in plane-polarized light containing wavelengths which cover the absorption-band of the substance, these will be more strongly absorbed when the electric vector of the polarized light vibrates parallel to the direction of the molecules or of their chromophore groups (e g systems of conjugated double bonds) than in other directions. In a crystal, or ordered fibre, the directions of varying pleochroic absorption, and hence of colour, may be related to principal crystallographic directions; in

uniaxial crystals and fibres, dichroism is related to two refractive indices, in biaxial crystals there are three refractive indices, so that trichroism may result (Bernal and Crowfoot, 1933; Morton, 1946; Bunn, 1946; Hartshorne and Stuart, 1950). In the case of haemoglobin, the four haem groups, consisting of iron atoms co-ordinated at the centres of essentially planar porphyrin molecules, are strongly light-absorbing; the exact form of the absorption in each derivative is influenced among other things by the protein part of the molecules. Perutz (1939, 1949) has analysed the dichroism of monoclinic crystals of horse methaemoglobin, finding that the absorption in the red and green spectral regions is far more intense for light vibrating with its electric vector parallel to the β and γ-crystal directions than to the α-direction. The *a* crystal plane contains *b* and *c* optic axes, and this would appear to be the plane containing the four haem groups, lying approximately parallel to each other, probably on the surface of the molecules by inference from chemical transformations of the crystals observed under the micro-spectroscope. Recent X-ray evidence has suggested that the polypeptide chains of the horse methaemoglobin molecule are α-helices arranged largely parallel to the *a* axis, to which the flat haem groups are nearly normal (Bragg, Howells and Perutz, 1952; Perutz, 1952; Crick, 1952), it seems that the helices turn corners after running for, perhaps, a quarter the length of the molecule. The weakly-absorbing α-direction at 16° to *a* axis is nearly parallel to the length of the molecule (Bragg and Perutz, 1952b). Further evidence for the existence of the α-helix polypeptide chain (Pauling, Corey and Branson, 1951; Pauling and Corey, 1951) in haemoglobin has been found by Perutz (1951) in the presence of an X-ray reflexion at 1 50 Å spacing, similar to the stronger reflexion in horse-hair, wool, muscle, in fibrous proteins and in poly(γ-benzyl L-glutamate).

Perutz (1952) has found further interesting implications of the haem pleochroism. The dichroic ratio of human reduced haemoglobin is very high (4 5), and as in monoclinic horse methaemoglobin, the direction of low absorption is nearly [100], and parallel to *a* axis When converted to HbO_2, HbCO or methaemoglobin in concentrated phosphate buffer, there is no

apparent change of crystal form but the dichroic ratio falls. When the experiment is performed in the opposite direction, the ratio rises, but the absorption maximum of the reduced haemoglobin formed is 150 Å on the long-wave side of its normal position; on solution of such a crystal, the normal maximum is regained. These changes may be related both to a weakening of haem-protein interaction and to a greater degree of parallelism of the four haem groups on reduction.

Pleochroic crystals may also exhibit anomalous dispersion of the birefringence when small and thin. This is due to the refractive index in the absorbing direction rising to a maximum on the long-wave side of the absorption peak, and reversing in sign on the short-wave side (see Wood, 1949). Sometimes, as with dyed protein and nucleic acid fibres, both phenomena may be utilized to detect the lie of the absorbing groups of the coloured molecules (White and Elmes, 1952). An interesting example involving reduced haemoglobin crystals will be described in connection with sickle-cells (Perutz and Mitchison, 1950).

The dichroism of monoclinic crystals of adult human reduced haemoglobin, crystallized from 2·4 M-Phosphate buffer at pH 6·7, is illustrated in Plate XXVI, Fig. 1.[1] Maximal absorption is in the direction of elongation of the crystals, and extinction is slightly oblique (about 3°). In Plate XXVI, Fig. 2, the use of the First order gypsum plate ('the sensitive tint') shows that the direction of highest refractive index is also in the direction of elongation. In solubility experiments on human reduced haemoglobin (see below), such crystals have frequently been found, and also forms with the needle orientation parallel to [100], maximal dichroic absorption and refractive index being at right angles to the elongation, as in the cases studied by X-rays. Fan-shaped assemblies of plates, elongated across the axis of the fan, are also found; maximal dichroic absorption and refractive index are across the fan, and a biaxial interference figure with crossed dispersion is given, indicating that the crystallographic axes and acute bisectrix almost coincide; the plates of such fans probably have {001} habit.

[1] The plates referred to in the course of this lecture will be found at the end of the book

Crystallography of Foetal Human Haemoglobins

Human foetal haemoglobin is crystallographically distinct from adult (Perrier and Jannelli, 1931; Haurowitz, 1935). Jope and O'Brien (1949) showed that foetal HbO_2, HbCO and methaemoglobin crystallized isomorphously like the adult protein, but belonged to a different crystal system from the latter—probably triclinic or monoclinic (Plate XXVIII, Fig. 8). As in the adult systems, the reduced foetal haemoglobin was crystallographically distinct, and possibly monoclinic. The difference between human adult and foetal haemoglobins has been confirmed by X-ray diffraction by Kendrew (1949).

Kendrew and Perutz (1948) have carried out detailed X-ray crystallographic analysis of adult and foetal *sheep haemoglobins.* The former is monoclinic (Space group C2), while two forms of foetal pigment have been characterized as orthorhombic ($B22_12$) and monoclinic ($P2_1$). These have presented some unusual difficulties of correlation between optical and X-ray diffraction data, but indicate considerable structural differences between adult and foetal forms (Bragg and Perutz, 1952b).

Solubility Studies on Adult and Foetal Human Haemoglobins

The effect of salts on the solubility of proteins has long been known Small concentrations of neutral salts may greatly increase the solubility, and also the stability The globulins, for *instance are insoluble in distilled water but freely soluble in* dilute salt solutions. In very dilute salt solution, the pH of minimum solubility is close to the iso-electric point of the protein, though in stronger solutions the pH value for minimum solubility may differ from the true iso-electric point In strong salt solutions, most proteins tend to precipitate—the salting-out effect. Strong phosphate buffers (Cohn, 1927, Green, 1933) have been used particularly for this purpose, the pH remaining constant, and also ammonium sulphate solutions. A single homogeneous protein will also show a definite, fixed solubility in a given strong salt solution at constant pH, temperature and pressure (Sørensen and Høyrup, 1917), and the application of

the phase rule to protein characterization has been followed particularly by Northrop and his school (Northrop, Kunitz and Herriott, 1939). Both the nature and charge of the ions influence the ease of precipitation of proteins (Hofmeister, 1887–8). Debye (1927) has explained the salting-out of proteins as due to their being squeezed out of solution by the salt ions attracting the polarizable water molecules around them. A full discussion of the physical chemistry of protein solubility will be found in the monographs of Cohn and Edsall (1943) and of Bull (1951)

Mellanby (1905) carried out early studies on the solubility of serum globulin, and found that in dilute solution, positive or negative ions with equal valencies were equally effective in increasing solubility, and that the efficiency of the ions of differing valency was directly proportional to the square of the valency. The ionic strength of electrolyte solutions was formulated by Lewis and Randall in 1921. This entered into the thermodynamic treatment by Debye and Huckel (1923) of interionic attraction; the latter workers employed as the expression for ionic strength, half the sum of all the terms obtained by multiplying the molar concentration of each ion by the square of its own valence—μ or $r/2 = 0{\cdot}5\ \Sigma\ C_iZ_i^2$. Proteins are dipolar ions (zwitterions), influencing the di-electric constant, and hence interionic attraction in electrolyte solutions. In 1925, Cohn showed that the solubility of haemoglobin and other proteins in concentrated ammonium sulphate solutions could be expressed in linear form by plotting log solubility as against ionic strength.

$$\log S = \beta' - K_s'(r/_2)$$

where S is solubility in g/l, K_s' the salting-out constant, $r/2$ ionic strength, β' the extrapolated intercept of the linear portion of the curve on the ordinate axis (log S). Haemoglobin solubility in various salt solutions is only linear over a range of high ionic strengths, and closely resembles cystine in behaviour (Cohn, 1936) The salting-out constant K_s' is independent of pH and temperature for a given salt and protein, whereas β' is markedly influenced and reflects the amphoteric properties of the protein. Green (1931) obtained series of parallel lines when log S was plotted against $r/2$ for horse HbO_2 and HbCO at

varying temperatures, and similarly at varying pH; at 0° C. horse HbCO is about 25 times more soluble than at 25° C. at high r/2, but at low salt concentrations, solubility increases with temperature.

Green, Cohn and Blanchard (1935) studied the solubility of human and horse HbCO in concentrated phosphate buffers, and found such marked differences as to distinguish the two proteins completely. Human HbCO was far more soluble, and both β' and K_s' were greater. Jope and O'Brien (1949) have carried out solubility studies of the type described on human foetal and adult haemoglobins. HbCO, crystallized from adult blood, behaved as a single component when increasing amounts of the solid phase were mixed with 2M.phosphate at pH 6·7. At this pH, adult HbO_2, HbCO and methaemoglobin had rather similar solubilities and obeyed the Cohn equation over the range of r/2, 4 to 6. On the other hand, reduced haemoglobin was far less soluble; in the range r/2, 3·5 to 4·5, solubility was linear. Amorphous HbCO was more soluble than crystalline, as found for other species by Roche, Derrien and Moutte (1941), and higher solubilities were obtained by dissolving up HbCO than by salting out.[1]

Jope and O'Brien (1949) observed a difference between reduced Hb and the isomorphous derivatives, HbO_2, HbCO and methaemoglobin, also in the effect of temperature. Hb did not vary in solubility with temperature, whereas the latter three had temperature-solubility curves with minima at about 20° C. Distinction between foetal and adult HbO_2 was also apparent in this respect, the solubility of foetal HbO_2 increased linearly with temperature from 0° to 25° C., resembling the behaviour of many proteins in dilute salt solutions. Foetal reduced Hb was only slightly more soluble than adult, and similarly was unaffected by temperature.

The distinctions established by solubility measurements have recently acquired great practical importance, particularly with regard to the haemoglobin abnormality in sickle-cell anaemia.

[1] A recent review by J. Roche and Y. Derrien (Sang, 24, 97, 1953) deals with further aspects of the solubilities of haemoglobins.

Differences in the Amino-acid Composition of Foetal and Adult Haemoglobins

Studies on adult and foetal ox haemoglobins (Vickery, 1944) indicated differences in the amino-acid composition, but the most revealing work is that of Porter and Sanger (1948). Using the 1 : 2 : 4-fluorodinitrobenzene reagent (Sanger, 1945) for estimation of free amino-groups in proteins, these workers found that human adult haemoglobin contained 5 terminal valine residues, and 43 free ε-amino-groups of lysine, whereas the foetal protein contained 47 free lysine ε-amino-groups and an average of 2 6 terminal valine residues. A molecular weight of 66,000 was assumed in each case; Jope and O'Brien (1949) mention that ultra-centrifugal and osmotic pressure studies by Gutfreund suggest similar molecular weight for foetal and adult human haemoglobin, and a tendency to dissociate on dilution is similar in the two. Andersch, Wilson and Menten (1944) suggested a molecular weight of half the adult value however, from the sedimentation constant alone.

Beaven, Hoch and Holiday (1951) point out that as even a 30-week foetal haemoglobin sample would contain some adult pigment, the results of Porter and Sanger probably indicate the presence of two valyl end groups in the foetal molecule. The number of free α-amino-groups indicates the number of open peptide chains in the molecule, and it is probable that this is also equal to the total number of chains present. The finding of five terminal valyl groups in adult human Hb and six in the closely-similar horse haemoglobin molecule is interesting.

The Immunological Specificity of Foetal and Adult Human Haemoglobin

The antigenic specificity of the globin in haemoglobins from different animals has been studied extensively (Heidelberger and Landsteiner, 1923; Hetkoen and Schulhof, 1923; Hetkoen and Boor, 1931), and the difference between human foetal and adult forms was demonstrated clearly by Darrow, Nowakovsky and Austin (1940). Porter and Sanger (1948) consider that there is a relationship between the terminal amino-acid residues and immunological specificity

ALKALI DENATURATION OF FOETAL AND ADULT HAEMOGLOBIN

When haemoglobin solution is made strongly alkaline with sodium hydroxide, the adult pigment is rapidly denatured and the brown colour of alkaline haematin appears. As long ago as 1866, Körber observed that foetal blood was much more resistant to this denaturation, the red colour persisting for an hour or so. This difference has been studied by numerous workers (von Kruger, 1888; Bischoff, 1926; Haurowitz, 1929, 1930; Trought, 1932). Brinkman and co-workers (Brinkman, Wildschut and Wittermans, 1934; Brinkman and Jonxis, 1935, 1937) have developed methods for following the reaction photometrically at pH 12·7 through the increase in light absorption in the red as the alkaline haematin is formed, the results are expressed graphically as a first-order reaction. Mixtures of foetal and adult haemoglobin give two-component curves, and the log percentage of the slower foetal component is found by extrapolating back to the ordinate. It is also claimed that at pH 11·7 adult haemoglobin denatures slowly, and may show evidence of two components with different resistance.

Jonxis (1949) found more foetal haemoglobin in the blood of premature infants than of full-term children, but in both a proportion of adult form was present. The foetal form diminished steadily after birth and at 4½ months had disappeared. He (Jonxis, 1948, 1949) also claimed that in erythroblastosis foetalis due to rhesus antagonism, the cells containing foetal haemoglobin disappeared more rapidly, through selective haemolysis. This view was attacked by Baar (1948), as he and Mollison (1943) found that transfused cells were destroyed as readily as the erythroblastotic infant's own cells. By denaturation studies, Ponder and Levine (1949) found no difference in the rate of disappearance of foetal Hb in normal and erythroblastotic infants.

It has been shown that alkali denaturation of methaemoglobin proceeds more rapidly than that of HbO_2 (Lemberg and Legge, 1949; Fig. 3). At a given pH, HbO_2 is denatured more slowly at high ionic strength of salts (Lewis, 1926); at low ionic strength, the denaturation rate is proportional to $[OH^-]$ concentration

(Haurowitz, 1931). The absorbtion of the alkaline haematin solution is far greater than the corresponding amount of HbO_2 in the spectral region 600–650 mμ (Donaldson, Sisson, King, Wootton and Macfarlane, 1951). Keilin (1926) believed that the denatured globin and alkaline haematin would be free in solution, but Anson and Mirsky (1930) and Drabkin and Austin

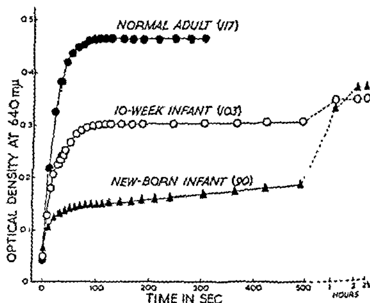


FIG 3 Denaturation of HbO_2 in 0 04 N.NaOH at 20° C.

(1935) thought a conjugated protein of ferri- or ferro-haemochrome type was formed. On spectrophotometric grounds, King and Delory (1945) considered that a conjugated protein of oxidized haematin and denatured globin was formed. It should also be remembered that under such conditions as low temperature, the alkali denaturation of haemoglobin may be reversible (Hill, 1925).

In Fig. 3, the denaturation curves of HbO_2 of a normal adult, 10-week infant and new-born infant are plotted as rise of density at 640 mμ (spectral red filter) against reaction time in seconds. 0 2 ml. of N.NaOH is rapidly stirred into 4·8 ml. of HbO_2 solution, of initial D_{640} about 0 08, in the 1 cm. photo-

electric colorimeter cell, and the density recorded at 10-second intervals. Whereas denaturation of the adult protein is complete in less than 100 seconds, the infant samples give a rapid initial rise in density, followed by a slow increase which is complete in 2–2½ hours. The curves are far more informative when expressed in the form of a first order, monomolecular reaction (see Glasstone, 1947). In the course of the reaction, the $[OH^-]$ remains substantially constant, so that the plot of the log of the concentration of unchanged molecules of HbO_2 against time is a straight line of slope

$$-\frac{2{\cdot}303}{\text{Reaction Constant K.}}$$

It can be shown that any time t during the reaction, the concentration of unchanged HbO_2 is proportional to

$$\frac{D_\alpha - Dt}{D_\alpha - Do}$$

where:

Do = Initial Optical Density at 640 mμ,
$D_\alpha = D_{640}$ at completion of reaction,
$Dt = D_{640}$ at time t.

When the data from Fig. 3 are expressed in this form, the lines shown in Fig. 4 are obtained. The adult HbO_2 yields a straight line of steep slope and reaction half-time 12 sec. (50 per cent of unchanged HbO_2 present) The 10-week infant HbO_2 yields a two-component curve—an initial steep slope due to the adult HbO_2 present, and a slow component with a t½ of 1150 sec., due to foetal HbO_2. By extrapolation, the percentage of foetal HbO_2 is found to be 16 per cent of the total. In the new-born infant, the foetal HbO_2 is 73 per cent of the total.

Reaction rates are extremely sensitive to temperature (Arrhenius, 1889) and precise thermostatic control presents considerable difficulty. In the experiments described here, the temperature was 20° ± 1° C. A joint line was derived by the method of least squares for the denaturation of 16 samples of normal adult HbO_2 in 0·04 N.NaOH, giving t½ 11·3 seconds, S.D. 2·3 seconds. No evidence has been found for the presence of a second component in normal adult haemoglobin, with different

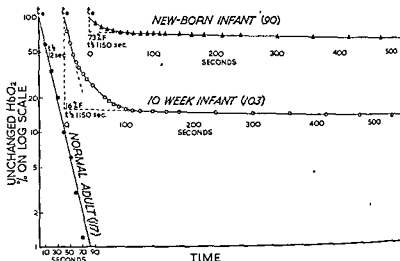


FIG 4 Denaturation of HbO_2 in 0 04 N.NaOH at 20° C. expressed as first-order reaction.

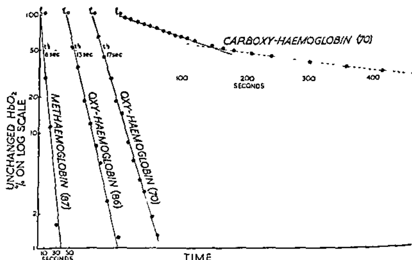


FIG. 5. Denaturation of HbO_2 in 0 04 N.NaOH at 20° C. compared with MET-Hb and CO-Hb.

denaturation rate at pH 12·8. In young infants, small amounts of foetal haemoglobin are readily detected, and in artificial mixtures of adult and foetal HbO_2, as little as 2·5 per cent of the latter could be detected, but not accurately measured. The denaturation of methaemoglobin and HbCO is compared with that of HbO_2 in Fig. 5. Methaemoglobin is more rapidly denatured, but in the carboxy-compound, the stable CO—haem appears to protect the protein and a slow denaturation occurs. The curve is complex, more probably due to secondary reactions than to the presence of a second component; adult HbCO was found to be homogeneous in phase-rule studies (Jope and O'Brien, 1949).

Monomolecular Films of Adult and Foetal Haemoglobins

Proteins, like many polar compounds, may be spread on the surface of water or dilute salt solutions to form monomolecular layers, the structure of which can give information on molecular size and form (see Sobotka, 1947). Stable monomolecular films of maximal area form in the neighbourhood of the iso-electric point of proteins, and Brinkman and Jonxis (1935, 1937) have applied the method of Garter and Grendel (1926) to the study of haemoglobin. The area of spreading on the surface of M/300 phosphate buffers over the range pH 6 to 8 was measured, and plotted as area of film per mg of protein after 1 minute, against pH. Foetal and adult haemoglobins differed, the foetal form spreading more slowly than the adult, and maximal area being achieved at pH 6 7 instead of 6 9. These differences are reversed in foetal and adult haemoglobins of the cow, sheep and goat, as are the resistances to alkali (Jonxis, 1949).

Electrophoretic Mobility of Foetal and Adult Haemoglobins

Landsteiner, Longsworth and van der Scheer (1938) applied the electrophoretic method of Tiselius (1937) to haemoglobin of various animals, and found small but characteristic mobility differences. Andersch, Wilson and Menten (1944) studied HbCO from adults, and blood of young infants. The Longsworth patterns showed clear distinction between foetal and

adult HbCO, with resolution of the two components in the blood of infants.

Beaven, Hoch and Holiday (1951) studied adult and foetal oxy-haemoglobins in phosphate buffers at pH 7·1 and 8, r/2 0·05 (Hoch, 1949). Adult HbO_2 had slightly greater anodic mobility in both buffers. Under the conditions of marked boundary anomaly resulting from low ratio of ionic strength to protein concentration (Hoch, 1950), the haemoglobin from 20-week foetuses was found to contain a proportion (at least 6 per cent) of faster component, and this increased steadily throughout intra-uterine life and early infancy, replacing the foetal haemoglobin at about four months.

Electrophoresis of haemoglobins will be discussed further in connection with sickle-cell Hb.

Spectrophotometry of Foetal and Adult Human Haemoglobins

The strong absorption bands of haemoglobin derivatives in the visible are mainly determined by the haem groups, though the position of the maximum of the intense Soret band (Soret, 1878) in the violet, which is characteristic of the intact porphyrin ring, is influenced by the protein. However, the characteristic position for each derivative is constant for a wide range of mammalian, avian and amphibian haemoglobins, and in adult and foetal human haemoglobin (Jope, 1949). The Soret band of the expected intensity is present in the spectrum of the intact erythrocyte, if optical precautions to prevent light scatter are taken (Rubinstein and Ravikovich, 1946; Jope, 1949).

In 1937, Holiday described a sensitive means for resolving the fine structure of the ultra-violet absorption of proteins, which is due to their content of aromatic amino-acids and tryptophan. The method consists in moving a photographic plate vertically past the horizontal continuous spectral image of the slit of the spectrograph, at a rate which increases logarithmically as the distance travelled. By this means, Jope (1949) found that although the band [illegible] in the free amino-acid and in [illegible] roteins it is shifted to a longer [illegible] In human

foetal haemoglobin on the other hand, the position of the 'tryptophan notch' approached 289–8 mμ, both in solution and in the erythrocyte. The notch is resolved in foetal haemoglobin, whereas the adult haemoglobin has an unresolved inflection at 291 mμ.

Subsequently (Beaven, Holiday and Jope, 1950; Beaven, Hoch and Holiday, 1951), it has been found that at various stages of foetal and early infant life, the tryptophan notch occupies an intermediate position dependent on the proportions of adult to foetal haemoglobin. The relationship is not linear, as the better-resolved foetal structure *dominates* the position, and the greater part of the shift occurs in the range 50–90 per cent adult Hb. Less than about 10 per cent of foetal Hb cannot be detected.

TABLE 1. Foetal Haemoglobin in New-born and Young Infants

Normal	U.V. spectrum		Alkaline denaturation	
	Tryptophan Notch	% foetal Hb	$t_{\frac{1}{2}}$ sec.	% foetal Hb
Full-term infants:				
No. 32	2901 Å	60	480	68
72	2900	70	750	78
56	2899	80	1025	70
90	2899	80	1150	73
10-week infant:				
No 103	2908	10–15	1150	16
Haemolytic anaemia				
3-months infant.				
No 54	2901	60	600	40
4-months infant				
No 6	2910 {2909, 2911 ± 1}	—	176	72

The position of the tryptophan notch in a series of new-born and young infants is shown in Table 1, and the percentage of foetal Hb deduced by this means is compared with the value obtained by alkali denaturation (Beaven, Holiday and White, 1953). Fair agreement is obtained, though the latter method is more sensitive for the detection of traces of foetal Hb. The two

phenomena are unlikely to be related, though both reflect aspects of the fine structural differences distinguishing adult and foetal human haemoglobins. On a comparative scale, quite complex

(Jonxis, 1949), but the tryptophan notch has the 291 mμ position in both (Jope, 1949).

The Oxygen Dissociation Curve for Foetal Blood

A number of investigators have established differences between the oxygen dissociation curves for foetal and adult blood, and also for the maternal blood in pregnancy. Leibson, Likhnitzky and Sax (1936) found that the curve for the blood of normal women varied within narrow limits whilst the maternal blood in

women. Essentially similar results have been found by several investigators (Haselhurst and Stromberger, 1931, Eastman, Geiling and De Lawder, 1933; Darling, Smith, Asmussen and Cohen, 1941), and the question is discussed in detail by Barcroft (1946).

It is generally considered that the differences in behaviour of the foetal corpuscles is best accounted for by the presence of a different haemoglobin, but the situation is complicated by possible differences in ionic environment within the corpuscle, permeability and blood pH. Haurowitz (1935) found that in dilute solution, human foetal haemoglobin was unique in having a lower oxygen affinity than the maternal haemoglobin, and this reversal of the behaviour found in the corpuscles has been confirmed by Hill (see Barcroft, 1946), and by McCarthy (1943) for strong solutions of the haemoglobin. McCarthy (1943) found that the maternal corpuscles markedly decreased the oxygen affinity of the adult Hb, whereas the foetal cells had little effect on the foetal Hb, thus accounting for the position when the whole bloods were compared.

An interesting question, which has been little explored, is the

ease of oxidation of foetal haemoglobin to methaemoglobin. In blood of normal adults, methaemoglobin, or other pigment without oxygen-carrying power, is present only to a very small extent (Ramsay, 1949; Wootton, 1950). Künzer and Savelsberg (1951) found 2·12 per cent of methaemoglobin in the blood of suckling infants in the first three months, as against 1 per cent in the blood of all other age groups; cord blood samples contained only 1 per cent. They suggest that this relates to low activity of the glycolytic enzyme systems which normally prevent by reduction the intra-corpuscular accumulation of methaemoglobin (see Barcroft, Gibson and Harrison, 1949). Well water containing nitrates may cause methaemoglobinaemia in infants (Comly, 1945), and Cornblath and Hartmann (1948) considered that nitrite formed by reduction by microorganisms in the gastro-intestinal tract was responsible. Lecks (1950) obtained some evidence that more ready oxidation of foetal haemoglobin by nitrites also played a part. The frequent occurrence of heavy methaemoglobinaemia in premature infants treated with sulphamezathine (Mann and O'Brien, 1951) may also indicate this. Studies are needed on systems free of enzymes, or in the presence of inhibitors.

Clinical Significance of Foetal Haemoglobin

It is usual to point to the different sites of erythrocyte formation in foetus and adult in explaining the two types of haemoglobin. The liver as opposed to the marrow is often quoted as the principal site of origin for foetal erythrocytes (see Barcroft, 1946; and Drabkin, 1951) However, foetal haemopoiesis is a complicated process, passing through several morphologically distinct phases, and widely distributed in many organs and tissues (Gilmour, 1941). The liver is the chief site in mid-foetal life, but skeletal sites appear in the clavicle of the 43 mm. embryo (Gilmour, 1941), and the marrow progressively assumes the principal role from the fifth month of intra-uterine life onwards, and is the exclusive site from 15–21 days after birth (Gilmour, 1941).

The spectrophotometric data of Beaven, Hoch and Holiday (1951) indicate a very gradual change-over from foetal to adult-

type haemoglobin formation, commencing not later than the middle of intra-uterine life. Even haemoglobin from a 20-weeks foetus contained 6–7 per cent adult-type Hb, and at term about 20 per cent was present. It is possible that the foetal-type globin is an expression of the general foetal anabolism rather than the specific site of origin. These workers found that no more than 10 per cent of foetal haemoglobin persists 4 months after birth, and this would be compatible with the morphological observations of Gilmour (1941), that erythropoiesis is entirely intramedullary from 15–21 days after birth, but also indicates that only adult haemoglobin is formed after delivery. Other workers suggest that foetal haemoglobin disappears in the first few months of life (Trought, 1932; Haurowitz, 1935), but Singer, Chernoff and Singer (1951) claim to have detected a small percentage persisting as long as two years after birth in healthy infants, using a sensitive denaturation test. It is clear that further correlation is required between the morphology of haemopoiesis during intra-uterine life and infancy, and the detection of foetal haemoglobin by a range of physical methods.

It has frequently been considered that foetal haemoglobin may persist after birth, or reappear, in various pathological conditions (see Barcroft, 1934) or modified environmental conditions (e.g. in dwellers at high altitudes; Rubowitz, 1933). In pernicious anaemia, the similarities between megaloblastic haemopoiesis and the earliest, primitive generation of nucleated red cells in pre-somite embryos, and also the part played by the liver, have prompted several investigators to look for foetal haemoglobin. Schenk (1930) thought alkali-resistant haemo-

cells did not disappear after treatment by purified liver extracts, folic acid or vitamin B12. However, Trought (1931) and Haurowitz (1935) have found the rapid denaturation characteristic of normal adult Hb, and Jope and O'Brien (1949) prepared orthorhombic crystals of normal adult HbO_2 from pernicious anaemia blood.

Liquori (1951) investigated the blood in thalassaemia major

(Cooley's anaemia; Cooley and Lee, 1925), and obtained unequivocal evidence for the presence of a mixture of adult and foetal haemoglobins. This was indicated by slow alkali-denaturation rate, and an intermediate position of the tryptophan fine-structure feature on ultra-violet spectrophotometry (Holiday, 1937; Jope, 1949; Beaven *et al.*, 1951), and crystals of HbO_2 were prepared with the morphology and higher solubility characteristic of the foetal pigment. By measuring the slope of the tryptophan fine-structure feature at two selected wavelengths in the ultra-violet, Rich (1952) has consistently found large amounts of foetal haemoglobin in the blood of children with thalassaemia major, whereas the parents with thalassaemia minor or the trait have the adult pigment. Genetically, Cooley's anaemia appears to be a homozygous condition inherited from heterozygous parents (Moncrieff and Whitby, 1934; Valentine and Neel, 1944), and a fundamental defect in the ability to synthesize haemoglobin occurs, the presence of foetal Hb being the result rather than the cause of this. In Table 2, the presence of foetal haemoglobin is shown in the blood of a Cypriot child with Cooley's anaemia, whilst the parents have the adult pigment (Beaven, Holiday, Johnson and White, 1953).

TABLE 2. Mediterranean Anaemia

	U.V. spectrum		*Alkaline denaturation*	
	Tryptophan notch	% foetal Hb	t½ sec.	% foetal Hb
Father, trait	2910	—		
Mother, trait	2910	—		
Daughter, aet 3 Thalassaemia major	2907	20	113	12

By measuring the percentage alkali denaturation in one minute, and by a fractional denaturation method, Singer, Chernoff and Singer (1951) have detected foetal haemoglobin in Cooley's anaemia, in the trait in small amounts, and also they

TABLE 3. Anaemias with Adult-type Haemoglobin

	Adult-type U.V. spectrum	Adult alkaline denaturation pattern
Pernicious Anaemia	7	3
Acute Leukaemia	3	2
Myelosclerosis	1	1
Severe Hypochromic Anaemia	1	
Aplastic Anaemia	4	
Total Obs.	16/16	6/6

TABLE 4 Familial Haemolytic Anaemias with Adult-type Haemoglobin

	Adult-type U.V. spectrum	Adult alkaline denaturation pattern
Hereditary Spherocytic Anaemia	17	7
Hereditary Elliptocytosis	4	4
Paroxysmal Nocturnal Haemoglobinuria (Not familial)	1	1
Atypical Congenital Haemolytic Anaemias	4	1
Total	26/26	13/13

believe that small amounts of alkali-resistant pigment may occur in a variety of haematological disorders, including some cases of acute and chronic leukaemias, pernicious anaemia, leuco-erythroblastic anaemia due to secondary carcinomatosis of the marrow, chronic aregenerative anaemia, and hereditary spherocytosis; haemorrhagic and iron-deficiency anaemias have normal adult pigments only.

A wide variety of blood disorders has been examined by the previously described alkali-denaturation method, and by U.V. spectrophotometry (Beaven, Holiday and White, 1953), and the results are given in Tables 3 and 4. A series of haemolytic anaemias was studied, including the cases of atypical haemolytic anaemias recently published by Dacie, Mollison, Richardson, Selwyn and Shapiro (1953). Despite the frequent occurrence of profoundly disturbed haemopoiesis, including the existence of extra-medullary blood formation, the haemoglobin in all these cases appears to be predominantly of the adult type.

The evidence for occurrence of foetal haemoglobin in sickle-cell anaemia will be considered later.

Sickle-cell Haemoglobin

In 1949, Pauling, Itano, Singer and Wells announced that the haemoglobin in sickle-cell anaemia behaves differently to normal adult haemoglobin in the Tiselius electrophoresis apparatus. At pH 6·9 in phosphate buffer of $\Gamma/2$ 0·1, sickle-cell HbCO moved as a positive ion, the normal protein as a negative ion, and the iso-electric points were found to be pH 7·09 and 6·87 respectively. This small but significant difference corresponds to 2–4 more positive charges per sickle-cell haemoglobin molecule. In sicklaemia (the non-anaemic carrier trait) the Longsworth diagram indicated the presence of a mixture of normal and sickle-cell haemoglobins in about 60 : 40 ratio Wells and Itano (1951) found that this ratio of the two haemoglobins was usual in sicklaemia, but that cases of sickle-cell anaemia might also contain from 5 to 20 per cent of electrophoretically normal haemoglobin; under the conditions employed, normal adult and foetal haemoglobin would be indistinguishable.

Sickled cells are birefringent (Sherman, 1940), and Perutz

and Mitchison (1950) were able to show that they actually contain crystalline reduced haemoglobin, with optical properties corresponding to those of small needles and plates of reduced haemoglobin crystallized from buffer solutions. The extinction directions, pleochroism and anomalous dispersion of the birefringence all corresponded. They showed further that the characteristic of sickle-cell haemoglobin that causes it to crystallize in the cells is its extremely low solubility in the reduced state (about one-tenth that of normal reduced Hb), whereas the HbO_2 and HbCO compounds have the same solubility as normal—the sickle-cell reduced Hb would therefore be about 100 times less soluble than the oxy-compound. Later, Perutz, Liquori and Eirich (1951) showed that the reduced sickle-cell Hb has a far lower solubility than normal, not only when crystals are dissolved in buffers of high r/2, but also under more physiological conditions. It has further been found (Harris, 1950) that solutions of sickle-cell haemoglobin gel and form tactoids when exposed to low oxygen tension, and that these closely resemble sickled cells under the microscope. Singer and Chernoff (1952) showed that only sickle-cell haemoglobin forms tactoids, but that haemoglobin solution from the sicklaemia trait would also gel if very concentrated, and after long exposure to CO_2 in the absence of O_2.

The unique character of sickle-cell haemoglobin has been fully established, but the molecular differences from the normal protein are very fine. In an X-ray crystallographic study, Perutz, Liquori and Eirich (1951) found that normal and sickle-cell reduced haemoglobins are identical in this respect, and that sickle-cell HbO_2 also crystallized in tetragonal ($P4_12_1$) and orthorhombic ($P2_12_12_1$) forms identical with their normal counterparts. In addition, however, orthorhombic crystals of sickle-cell HbO_2 were prepared, with unique characters—space group $P2_12_12$, a large unit cell of wet volume 1150,000 $Å^3$ containing 8 molecules, and the molecular weight of the asymmetric units 136,000. This form may be a dimer (Drabkin, 1951).

Schroeder, Kay and Wells (1950) quantitatively analysed normal and sickle-cell haemoglobins from negroes for 17 amino-

acids and ammonia, but found only trifling differences. A similar minimum molecular weight for each was inferred, the isoleucine amounting to 1 residue for a M.W. of about 67,000. Havinga

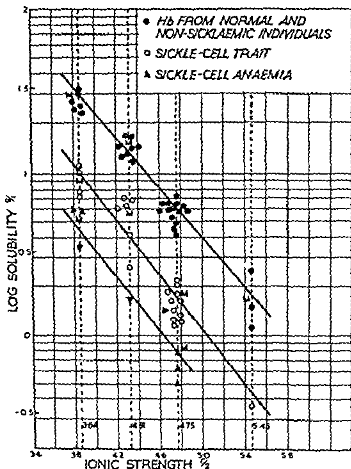


Fig. 7. Solubility of reduced haemoglobin in pH 6·7 phosphate buffer of varying ionic strength.

(1953) applied the end-group assay method of Sanger (1945), finding similar numbers of valyl end groups in normal and sickle-cell globins and intact haemoglobins, though the sickle-cell haemoglobin reacted with dinitrofluorobenzene more

rapidly than the normal. Havinga found that sickle-cell HbCO was more difficult to split into haem and globin with acid acetone, possibly indicating a difference in the regions of haem attachment to the protein. This is of interest when it is recalled that the hydrophobic character of the sickle-cell Hb only becomes apparent on reduction, when haem-protein interaction may weaken (Perutz, 1952). The soluble globins separated from normal and sickle-cell haemoglobins preserve characteristic differences in electrophoretic mobility unless vigorously denatured (Havinga and Itano, 1953).

The results of solubility studies on four cases of sickle-cell anaemia and nine sickle-cell traits (sicklaemia) are shown in Plate XXVII, Fig. 6 and in Fig. 7 and compared with haemoglobin from normal individuals, or anaemias not of sickling type. The experiments were set up in pH 6·7 phosphate buffers of varying ionic strength, using ferrous citrate as reducing agent, sealed in an atmosphere of nitrogen, and equilibrated with gentle mixing for five days at 22° C. Lines are drawn for the three groups through the group mid-points in each strength of buffer, and the linear Cohn relation between log solubilities and Γ/2 appears to be followed over the range Γ/2 3·84–5·45. The sickle-cell Hb has one-sixth to one-eighth the solubility of normal reduced Hb, and the mean solubility for the trait group is about twice this. The trait samples were from East Africans, and were kindly provided by Dr. H. Lehmann. Blood from a Gambian with ancylostoma anaemia, which did not sickle, contained haemoglobin with normal solubility.

The movement of normal and sickle-cell haemoglobins by electrophoresis on paper is shown in Plate XXVIII, Fig. 8. Spaet (1953) has also used this method for demonstrating the slight difference in mobilities between human haemoglobins.

Foetal Haemoglobin in Sickle-cell Anaemia

Using sensitive alkali-denaturation techniques, Singer, Chernoff and Singer (1951) found that 2–24 per cent of resistant haemoglobin frequently occurred in the erythrocytes of sickle-cell anaemia, but not in the trait. Combining this test with the test of tactoid formation for sickle-cell haemoglobin, Singer and

Chernoff (1952) recognized normal adult Hb(N) and sickle-cell Hb(S) in the trait cells, and in sickle-cell anaemia, the haemoglobin S might be accompanied by alkali-resistant, foetal-type Hb(F).

The presence of foetal haemoglobin in sickle-cell anaemia has also been established by Itano (1953 a and b) by the alkali-denaturation and spectrophotometric methods, and by solubility measurements. The presence of foetal haemoglobin in the sickle-cell increases the net solubility (Itano, 1953 a and b), which is advantageous to the individual. There is also some evidence that sickling is difficult to elicit in the young infants of sicklaemic parents, possibly due to a persistence of foetal haemoglobin (Watson, 1948)

In our own work, we have encountered foetal Hb in one case (*vide infra*) of sickle-cell anaemia out of six by alkali-denaturation and U.V. spectrophotometry, and none in nine traits. It is now known that foetal Hb may accompany sickle-cell Hb in conditions other than pure sickle-cell anaemia.

The Genetics of Sickle-cell Anaemia

Neel (1951, 1952) has postulated that the sickling phenomenon is due to a gene, which in single dose produces the trait, in double dose, the homozygous state with development of sickle-cell anaemia. This is well supported by the chemical evidence for a mixture of sickle-cell Hb and slight excess of normal Hb in the trait cells, but an excess of sickle-Hb, with perhaps some foetal Hb, in the anaemia. The theory requires that both parents should carry the gene, and their blood sickle. This is largely true, but exceptions occur.[1] In some cases, it appears that one parent has the sickle trait, the other the Thalassaemia trait (Cooley or Mediterranean trait), and that although the genes for the two conditions are separate, they may interact to produce a haemolytic syndrome with features of both sickle and Mediterranean anaemias in some of the offspring (Powell, Rodarte and Neel, 1950, Sturgeon, Itano and Valentine, 1952).

[1] Since this lecture was prepared, Singer and Fisher ((1953) *Blood*, 8, 270) have shown that in some anomalous cases, the erythrocytes of the non-sickling parent may contain no more than a trace of sickle-cell Hb, detectable only by careful electrophoretic analysis

This condition was first described in Sicilian and Southern Italian families by Silvestroni and Bianco (1952), and called micro-drepanocytic disease.

Haemoglobins in Haemolytic Anaemia due to Sickle-cell: Thalassaemia Interaction

Itano (1953 a and b) has demonstrated that either normal adult or foetal Hb may accompany the sickle-cell Hb in these cases.

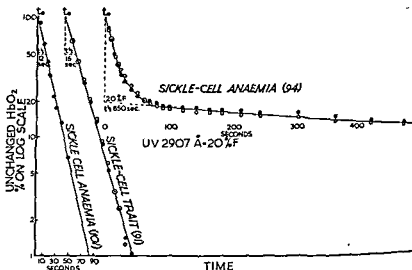


Fig. 9. Denaturation of HbO_2 in 0·04 N.NaOH at 20° C. Sickle-cell anaemias and trait.

We have studied blood from an Arab woman who may have the condition, which was kindly provided by Dr. N. A. F. Young from Kuwait. Sickle-cell Hb was demonstrated by the low solubility, similar to that of the sickle-traits studied; the presence of 20 per cent of foetal haemoglobin was indicated both by alkali-denaturation and U.V. spectrophotometry (Fig. 9). The mother's blood sickles, and the father appears to have thalassaemia minor.

A further family has been studied with Dr. Humble. The father and two sons have sickle-trait and haemoglobin of low solubility. The mother has the characteristics of thalassaemia minor. The third son is completely normal, and the

daughter is anaemic, sickles, and has haemoglobin of low solubility. No foetal Hb has been found in any member of this family.

To characterize fully the anaemias which may occur where one parent only carries the sickle-trait, electrophoretic analysis is necessary, since there is evidence for two further abnormal haemoglobins, controlled by genes which may interact with the sickling gene.

Other Abnormal Haemoglobins interacting with Sickle-cell Haemoglobin

Two further abnormal haemoglobins have been described by American workers, which may give rise to haemolytic anaemias of moderate severity. Itano and Neel (1950), and Kaplan, Zuelger and Neel (1951), described a new haemoglobin III, occurring in some American negro families, which had greater electrophoretic mobility than either normal or sickle-cell haemo-gl [illegible]

[illegible] trait; the other was a healthy, heterozygous carrier of the gene for the new haemoglobin, the erythrocytes containing normal adult Hb and Hb III. It is now known that some foetal Hb may accompany the sickle-cell Hb and Hb III in the haemolytic disease, and that the net solubility is not greatly reduced (Itano, 1953a)

[illegible] mixture of sickle-cell Hb and Hb d, and foetal Hb might be present (Itano, 1953 a and b).

It will be apparent from the foregoing that the situation is complicated, both genetically and biochemically, and it is possible that yet further patterns will emerge. The application of

the evidence for the classification of the four human haemoglobins at present known:

	Symbol (*Blood*, 8, 386, 1953)	(Itano, 1953a, b)	(Singer and Chernoff, 1952)
Normal Adult Hb	A	a	N
Normal Foetal Hb	F	f	F
Sickle-cell Hb	S	b	S
Haemoglobin III (second abnormal)	C	c	
Third abnormal Hb	D	d	

FURTHER ASPECTS OF SHAPE-ABNORMALITIES OF ERYTHROCYTES

In sickle-cell anaemia, crystallization of the abnormal haemoglobin when the oxygen-tension is reduced is largely responsible for sickling, and these cells are readily destroyed in the circulation. Other haemolytic anaemias may frequently exhibit abnormally-shaped erythrocytes, but contain no abnormal haemoglobin. An example was studied in a child of six, with probably a familial haemolytic anaemia, complicated by nephritis (Dacie *et al.*, 1953; Case 12). Triangular cells were present in the circulation, and preparations for sickling rapidly became birefringent due to formation of reduced Hb crystals within the cells (Plate XXIX, Fig. 10), though true sickling (Plate XXIX, Fig. 11) did not occur. The solubility of the haemoglobin was normal, and an abnormally permeable cell membrane may have been responsible.

Teitel-Bernard (1932) found that haemoglobin in normal human erythrocytes crystallized when the cells were slightly shrunken in hypertonic reducing media Isaacs (1950) observed sickle-like elongation of normal erythrocytes in very viscous media like glue, and we have found that normal cells become ellipsoidal, and their HbO_2 birefringent, in pH 6·7 M. phosphate buffer. The X-ray diffraction studies of Dervichian, Fournet and Guinier (1947) and Perutz (1948) indicate that the haemoglobin molecules in the normal erythrocyte can rotate freely, but are arranged in a close-packed lattice. This 'near order' is between the order of a solid crystal and the disorder of solution.

The role of the proteins and lipoids of stroma and surface membrane in maintaining the shape of the erythrocyte, and the

'near order' of the contained haemoglobin, has been stressed by Ponder (1948, 1950). Disproportion between stroma and the normal haemoglobin molecules may be responsible for the abnormally-shaped cells often seen in such conditions as hypochromic anaemias. Dervichian, Fournet, Guinier and Ponder (1952) have approached this problem in cells with abnormal haemoglobins by means of small-angle X-ray diffraction. To relate the structure of the erythrocyte adequately to all the various observed abnormalities of production and destruction, this aspect will have to be studied as closely as the possibility of abnormality in the haemoglobin itself.

CONCLUSION

The evidence for the existence of foetal and adult human haemoglobins, and for three genetically-controlled abnormal haemoglobins, is discussed and related to the clinical findings in various anaemic conditions.

The background of the physical chemistry of proteins has been stressed and elaborated at some length, as the various haemoglobins *can only* be recognized with certainty and differentiated by the application of precise methods. This is essential if their significance for the genetics, aetiology and clinical features of the anaemias concerned is to be properly assessed.

The unpublished investigations described in this lecture were carried out in co-operation with the M.R.C. Spectographic Unit, London Hospital, director Dr. E. R. Holiday, and Drs. G. H. Beavan and E. A. Johnson.

REFERENCES

AMANTEA, G. (1924) *Ber ges. Physiol* **23**, 167.
ANDERSCH, M A., WILSON, D A and MENTEN, M. L. (1944) *J. biol Chem.* **153**, 301.
ANSON, M. L and MIRSKY, A E (1930). *Physiol. Rev.* **10**, 506.
ARRHENIUS, S. (1889). *Z phys. Chem* **4**, 226.
BAAR, H S. (1948). *Nature, Lond.* **162**, 190
BARCROFT, H, GIBSON, Q. H. and HARRISON, D. C. (1949). *Haemoglobin. Barcroft Memorial Volume*, p. 223. Ed F J W. Roughton and J. C. Kendrew Butterworth, London
BARCROFT, J. (1934). *Features in the Architecture of Physiological Function.* Cambridge University Press.
BARCROFT, J (1946). *Researches on Pre-Natal Life Vol.* 1 Blackwell, Oxford.

BEAVEN, G. H., HOCH, H. and HOLIDAY, E. R. (1951). *Biochem. J.* 49, 374.

BEAVEN, G. H., HOLIDAY, E. R., JOHNSON, E. A. and WHITE, J. C (1953) Unpublished observations.

BEAVEN, G. H., HOLIDAY, E. R. and JOPE, E. M. (1950) *Disc. Faraday Soc* No. 9. 406.

BEAVEN, G. H., HOLIDAY, E. R. and WHITE, J. C. (1953). Unpublished observations.

BERNAL, J. D. and CROWFOOT, D. (1933). *Trans. Faraday Soc.* 29, 1032.

BISCHHOFF, H. (1926). *Z. ges. exp. Med.* 48, 161, 472.

BRAGG, W. L. (1952). *Proc phys. Soc. Lond.* B. 65, 833.

BRAGG, W. L., HOWELLS, E. R. and PERUTZ, M. F. (1952). *Acta cryst Camb.* 5, 136.

BRAGG, W. L. and PERUTZ, M. F. (1952a). *Acta cryst.* 5, 277

BRAGG, W. L and PERUTZ, M F (1952b). *Ibid.* 5, 323.

BRAGG, W. L and PERUTZ, M. F. (1952c). *Proc. roy. Soc.* A. 213, 425

BRINKMAN, R. and JONXIS, J. H. P. (1935). *J. Physiol.* 85, 117.

BRINKMAN, R. and JONXIS, J. H. P. (1937). *Ibid.* 87, 162.

BRINKMAN, R , WILDSCHUT, A. and WITTERMANS (1934). *J. Physiol.* 80, 377.

BULL, H. B. (1951). *Physical Biochemistry.* 2nd Ed. Wiley and Chapman and Hall, New York and London

BUNN, C. W. (1946). *Chemical Crystallography*, pp. 85–7, 285–7. Oxford University Press.

. . 173.

COHN, E. J and EDSALL, J T. (1943). *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions.* Reinhold, New York.

COMLY, H. H (1945). *J. Amer. med. Ass.* 129, 112.

COOLEY, T. B. and LEE, P. (1925). *Trans. Amer. pediat. Soc.* 37, 20.

CORNBLATH, M. and HARTMANN, A. F. (1948). *J Pediat.* 33, 421.

CRICK, F H C (1952) *Acta cryst. Camb.* 5, 381.

DACIE, J. V., MOLLISON, P. L., RICHARDSON, N , SELWYN, J. G. and SHAPIRO, L. (1953). *Quart J Med* N S 22, 79.

DARLING, R C , SMITH, C A , ASMUSSEN, E. and COHEN, F. M (1941). *J clin. Invest.* 20, 739.

DARROW, R. R , NOWAKOVSKY, S and AUSTIN, M H. (1940). *Arch Path.* 30, 873

DEBYE, P. (1927) *Z. phys. Chem.* 130, 56.

DEBYE, P. and HUCKEL, E. (1923). *Phys Z* 24, 185, 305, 334.

DERVICHIAN, D G., FOURNET, G , GUINIER, A and PONDER, E. (1947). *C.R. Acad. Sci., Paris.* 224, 1848

DERVICHIAN, D. G , FOURNET, G , GUINIER, A. and PONDER, E (1947). *Rev. Hemat.* 7, 567.

DONALDSON, R., SISSON, R. B., KING, E J., WOOTTON, I. D. P. and MACFARLANE, R. G. (1951). *Lancet*, i. 874.

DOUGLAS, C. G., HALDANE, J. S. and HALDANE, J. B. S. (1912). *J. Physiol.* **44**, 275.
DRABKIN, D. L. (1945). *Amer. J. med. Sci.* **209**, 268
DRABKIN, D. L. (1946). *J biol. Chem.* **164**, 703.
DRABKIN, D. L. (1951). *Physiol. Rev.* **31**, 345.
DRABKIN, D. L and AUSTIN, J. H. (1935). *J. biol. Chem.* **112**, 89
EASTMAN, N J, GEILING, E M. K. and DE LAWDER, A. M. (1933). *Bull. Johns Hopk. Hosp.* **53**, 246.
FUNKE, O (1852) *Atlas of physiological chemistry.* Supplement to Lehmann's *Physiological Chemistry.* Plate X, Figs. 1 and 6. London.
GARTER, E. and GRENDEL, F. (1926). *Trans. Faraday Soc* **22**, 477
GILMOUR, J. R. (1941). *J. Path. Bact.* **52**, 25.
GLASSTONE, S (1947). *Text Book of Physical Chemistry*, ch. 13. 2nd Ed. Van Nostrand and Macmillan, New York and London.
GREEN, A. A. (1931). *J. biol. Chem.* **93**, 495, 507, 517.
GREEN, A. A. (1933). *J Amer chem. Soc.* **55**, 2331.
GREEN, A. A., COHN, E J. and BLANCHARD, M (1935). *J. biol. Chem* **109** 631.
HARRIS, J. W. (1950). *Proc. Soc. exp. Biol. N.Y.* **75**, 197.
HARTSHORNE, N. H and STUART, A (1950) *Crystals and the Polarising Microscope*, pp 238-41. 2nd Ed. Arnold, London
HASELHURST, G. and STROMBERGER, K (1931). *Z. Gebürtsh. Gynak.* **100**, 48.
HAUROWITZ, F. (1929) *Hoppe Seyl Z.* **183**, 78
HAUROWITZ, F. (1930). *Ibid.* **186**, 141.
HAUROWITZ, F. (1931). *Ibid.* **194**, 98.
HAUROWITZ, F. (1935). *Ibid* **232**, 125
HAVINGA, E (1953) *Proc. nat. Acad. Sci. Wash* **39**, 59
HAVINGA, E and ITANO, H. A. (1953). *Ibid* **39**, 65.
HEIDELBERGER, M. and LANDSTEINER, K. (1923). *J. exp. Med.* **38**, 561.
HETKOEN, L. and BOOR, A. K (1931) *J. infect Dis* **49**, 29
HETKOEN, L. and SCHULHOF, K. (1923) *Ibid* **33**, 224.
HILL, R (1925). *Biochem. J* **19**, 341.
HOCH, H (1949). *Ibid.* **45**, 285.
HOCH, H. (1950). *Ibid.* **46**, 199.
HOFMEISTER, F. (1887-8). *Arch. exp. Path. Pharmak.* **24**, 247.
HOLIDAY, E. R. (1937) *J Sci. Instrum.* **14**, 166
HUNEFIELD, F. L. (1840). *Die Chemismus in derf thierischen Organisation*, p. 160. Leipzig.
ISAACS, R (1950) *Science.* **112**, 718.
ITANO, H A. (1951). *Proc. nat Acad. Sci. Wash.* **37**, 775.
ITANO, H. A (1953a). *Science* **117**, 89
ITANO, H. A. (1953b). *Fed Proc.* **12**, 224
ITANO, H A and NEEL, J. V. (1950) *Proc nat. Acad Sci Wash.* **36**, 613.
JONXIS, J. H. P (1948). *Nature, Lond* **161**, 850
JONXIS, J. H. P. (1949). *Haemoglobin Barcroft Memorial Volume*, p. 261. Ed. F. J. W. Roughton and J. C. Kendrew. Butterworth, London.

JOPE, E. M. (1949). *Haemoglobin: Barcroft Memorial Volume*, p. 205 Ed. F. J. W. Roughton and J. C. Kendrew. Butterworth, London.
JOPE, E. M., JOPE, H. M. and O'BRIEN, J. R. P. (1949). *Nature, Lond.* 164, 622.
JOPE, H. M. and O'BRIEN, J. R. P. (1949). *Haemoglobin. Barcroft Memorial Volume*. Ed. F. J. W. Roughton and J. C. Kendrew. Butterworth, London.
KAPLAN, E , ZUELGER, W. W. and NEEL, J. V. (1951). *Blood*. 6, 1240.
KEILIN, D. (1926). *Proc. roy. Soc.* B. **100**, 129.
KENDREW, J. C. (1949). *Endeavour*, **8**, 80.
KENDREW, J. C. and PERUTZ, M. F. (1948). *Proc. roy. Soc.* A. **194**, 375.
KENDREW, J. C. and PERUTZ, M. F. (1949). *Haemoglobin· Barcroft Memorial Volume*, p. 161. Ed. F. J. W. Roughton and J. C. Kendrew. Butterworth, London
KING, E. J. and DELORY, G E. (1945). *Biochem J*. **39**, 111.
KORBER, E. (1866). Doctoral Dissertation. Dorport.
KÜNZER, W. and SAVELSBERG, W. (1951). *Klin. Wschr.* **29**, 648.
LANDSTEINER, K. and HEIDELBERGER, M. (1923). *J. gen. Physiol.* 6, 131.
LANDSTEINER, K., LONGSWORTH, L. G. and VAN DER SCHEER, J. (1938) *Science*, **88**, 83.
LARSEN, G (1950). *Proc. 3rd Internat Congress: Internat. Soc. Haemat.* Paper No 76, p. 25 1951. Heinemann, London.
LECKS, H I. (1950). *Amer. J Dis. Child.* **79**, 117.
LECKS, H and WOLMAN, I J. (1950). *Amer. J. med. Sci.* **219**, 684
LEIBSON, R. G , LIKHNITZKY, I. I. and SAX, M. G. (1936). *J. Physiol.* 87, 97
LEMBERG, R. and LEGGE, J W. (1949). *Hematin Compounds and Bile Pigments*. New York and London, 1949.
LEWIS, G. N and RANDALL, M (1921). *J. Amer. chem Soc.* **43**, 1112.
LEWIS, P. S (1926) *Biochem. J* **20**, 984.
LIQUORI, A M (1951). *Nature, Lond.* **167**, 950.
MCCARTHY, E F (1943). *J. Physiol* **102**, 55
MAÇELA, I. and SELIŠKAR, A. (1925). *J Physiol.* **60**, 428.
MANN, T and O'BRIEN, D. (1951) *Lancet*, **i**, 1233
MELLANBY, J (1905). *Ibid* **33**, 338
MOLLISON, P. L. (1943). *Arch Dis Childh.* **18**, 161.
MONCRIEFF, A and WHITBY, L. E H. (1934). *Lancet*, **ii**, 648.
MORTON, T. H (1946) *J Soc. Dy. Col* **62**, 272.
MUNRO FOX, H (1945) *Nature, Lond* **156**, 18.
NEEL, J. V. (1951) *Blood*, **6**, 389
NEEL, J. V (1952) *Ibid* **7**, 467
NORTHROP, J H., KUNITZ, M. and HERRIOTT, R. M (1939) *Crystalline Enzymes*. Columbia University Press
PAULING, L. and COREY, R B. (1951). *Proc. nat. Acad Sci. Wash* **37**, 241, 261, 282, 729.
PAULING, L., COREY, R. B and BRANSON, H R. (1951). *Ibid* **37**, 205.

PAULING, L., ITANO, H. A., SINGER, S. J. and WELLS, I. C. (1949). *Science*, 110, 543.
[illegible]
[illegible] *Volume*, p. 135. Ed. F. J. W. Roughton and J. C. Kendrew. Butterworth, London.
PERUTZ, M. F. (1946). *Disc. Faraday Soc.* 42B, 187.
PERUTZ, M. F. (1948). *Nature, Lond.* 161, 204.
PERUTZ, M. F. (1952). *Symposium sur la Biochimie de l'Hematopoiese*, ii^e Congres International de Biochimie, p. 101. Paris.
PERUTZ, M. F., LIQUORI, A. M. and EIRICH, F. (1951). *Nature, Lond.* 167, 929.
PERUTZ, M. F., MITCHISON, J. M. (1950). *Ibid.* 166, 677.
PERUTZ, M. F., WEISZ, O. (1947). *Ibid.* 160, 786.
PONDER, E. (1948). *Haemolysis and Related Phenomena*, ch. 4. Grune and Stratton, New York.
PONDER, E. (1950). *Blood*, 5, 293.
PONDER, E. [illegible]
PORTER, R. [illegible]
POWELL, W [illegible]
RAMSAY, W [illegible]
Ed. F. J. [illegible]
REICHERT, E. T. and BROWN, A. P. [illegible]
Pub. No. 116. Carnegie Institute, Washington.
RICH, A. (1952). *Proc nat. Acad. Sci. Wash.* 38, 187.
ROCHE, J., DERRIEN, Y. and MOUTTE, M. (1941). *C.R. Soc. Biol. Paris* 135, 1235.
ROCHE, J. and DERRIEN, Y. (1953). *Sang*, 24, 97.
RUBINSTEIN, D. L. and RAVIKOVICH, H. M. (1946). *Nature, Lond.* 158, 952.
RUBOWITZ, M. (1933). *Biochem. Z* 266, 190.
SANGER, F. (1945). *Biochem. J.* 39, 507.
SCHENK, E. G. (1930). *Arch. exp. Path. Pharmak* 150, 160.
SCHROEDER, W. A., KAY, L. M. and WELLS, I. C. (1950). *J. biol. Chem.* 187, 221.
SHERMAN, I. J. (1940). *Bull. Johns Hopk. Hosp.* 67, 309
SILVESTRONI, E and BIANCO, I. (1952) *Blood*, 7, 429.
SINGER, K. and CHERNOFF, A I. (1952). *Ibid.* 7, 47.
SINGER, K., CHERNOFF, A. I. and SINGER, L. (1951). *Ibid.* 6, 413, 429.
SOBOTKA, H. (1947). *Medical Physics*, p. 763. Ed. O Glasser. Year Book Publishers, Chicago.
SORENSEN, S. P. L. and HOYRUP, M. (1917). *C.R Lab Carlsberg* 12, 213.
SORET, J. L. (1878). *Arch Sci. phys. nat.* 61, 322.
SPAET, T. H. (1953). *J. Lab. clin. Med.* 41, 157.
STURGEON, P., ITANO, H. A. and VALENTINE, W. N (1952) *Blood*, 7, 350.

TEITEL-BERNARD, A. (1932). *Arch. roum. Path. exp. Microbiol.* 5, 389
TISELIUS, A. (1937). *Trans. Faraday Soc.* 33, 524.
TROUGHT, H. (1932). *Arch. Dis. Childh.* 7, 259.
VALENTINE, N. W. and NEEL, J. V. (1944). *Arch. intern. Med.* 74, 185
VICKERY, H. B. (1944). *J. biol. Chem.* 156, 283.
VON KRUGER, F. (1888). *Z. Biol.* 24, 318.
WATSON, J. (1948). *Amer. J. med. Sci.* 215, 419.
WELLS, I. C. and ITANO, H. A. (1951). *J. biol. Chem.* 188, 65.
WHITE, J. C. and ELMES, P. C. (1952). *Nature, Lond.* 169, 151.
WOOD, R. W. (1949). *Physical Optics*, pp. 118–21, 469–74. 3rd. ed. Macmillan, New York.
WOOTTON, I. D. P. (1950). University of London, Ph.D. Thesis, Section 7.

XVII

The Life History of the Malaria Parasite

P. C. C. GARNHAM

MALARIA is said to be the greatest single killer of the human race. Vital statistics in the tropics, where the disease is commonest, are too unreliable to confirm the absolute truth of this statement but there can be little doubt that malaria is still—in spite of spectacular advances in control measures—one of the most important diseases of the world today. I expect most of you are familiar with the life history of the parasite—I intend in this lecture to mention briefly the well-known features, and shall spend most of the time in describing the latest discoveries regarding the tissue phase of the parasite, and shall indicate also what I think are the chief unsolved problems.

Malaria was known as a disease entity in the time of Hippocrates, and quotidian, benign tertian and quartan fevers with enlargement of the spleen were described. Except for the discovery of quinine, no major advances in our knowledge of the disease took place in the next 2,000 years; even at the end of the nineteenth century, malaria was thought to be transmitted by certain amoeboid cysts inhaled in the air of marshes, demonstrable in the dew or in nasal secretions of pigeons exposed for several nights in these regions.

[illegible]

time on, army doctors have continued to play a big part in filling in pieces in the jig-saw puzzle. In 1897, Captain Ronald Ross proved in India that *Plasmodium relictum*, the malaria

parasite of sparrows, was transmitted by mosquitoes, and in the following year, Grassi *et al.* (1898) in Rome elucidated the life cycle of *P. falciparum* of man in *Anopheles*. An important discovery was made by accident by Romanowsky, a Russian army surgeon, in 1891. He happened to mix old solutions of eosin and methylene blue, and in the latter, moulds had caused the formation of azur, on the presence of which the distinctive staining depends. Leishman's and Giemsa's stains are the result of this discovery. I am not going to describe the rest of the history in detail: Colonel James in 1931 stated that there must be a cryptic or occult stage of the parasite in the vertebrate host—Raffaele revealed it in Rome in 1934 in the bird parasite, *P. elongatum* ... i (and in *P. relictum* two years later) ... war, Colonel Short ... human forms in ... *...parum.*

Definition of a Malaria Parasite

New facts are being brought to light every year about the pigmented parasites of red blood corpuscles of vertebrates, and the old definition of a malaria parasite is now no longer appropriate. A provisional definition is as follows: The genus *Plasmodium* is comprised by pigment-producing amoeboid parasites with one habitat in the red blood corpuscles of the vertebrate host and another in its tissues. These two constitute the asexual cycle of the parasite. The sexual cycle takes place inside the insect, which is thus the definitive host. The asexual cycle is called schizogony and results in the formation of merozoites; the sexual is called sporogony and gives rise to sporozoites.

Cycle in the ...

The younge... ... enclosing a vacuole. Growth with a loo... is rapid an... the organism is amoeboid it assumes various shapes, throwing out pseudopodia. The parasite feeds on the haemoglobin of the red blood corpuscles and this is finally converted into minute grains of pigment which collect in the cytoplasm. The pigment is called haematin and consists of ferrihaemic acid. As the parasite grows, the vacuole disappears, and

the nucleus divides, giving rise eventually to the mature schizont with 4 to 32 merozoites, the number depending upon the actual species of *Plasmodium*. The schizont ruptures, the merozoites escape into the blood stream and invade new red blood corpuscles. This process continues, schizogony being repeated at regular and synchronous intervals—each time, the schizonts rupture, the patient experiences fever. Some species mature in 24 hours—like *P. knowlesi* of monkeys, where the fever is quotidian; others in 48 hours—like *P. falciparum* and *P. vivax*, giving rise to malignant and benign tertian malaria respectively; and others in 72 hours—like *P. malariae* which produces a quartan fever.

has fascinated me for a long time and I have tried to work on it a little with no success. Just as mysterious is why certain strains of parasites—particularly the newly discovered *P. berghei* of tree rats—gradually lose their ability to produce gametocytes, when passaged in the laboratory. Perhaps these problems are examples

are out of our grasp, and that our brains are not constructed to discover 'ultimate reasons'. Perhaps several of the problems I shall mention to you come into this category—but they still irritate!

Male and female gametocytes are produced inside erythrocytes and they circulate in the blood unchanged, waiting to be taken up by a mosquito, where they will undergo the next step in development. Before describing this, I shall just mention certain curious effects that the parasites produce on the red blood corpuscles they inhabit. They are interesting because firstly they are highly diagnostic of species and secondly because their nature is entirely unknown. These appearances can only be seen in Romanowsky-stained blood films. *P. vivax*—the parasite of benign tertian malaria—gives rise to a large number of

the youngest parasites, the oldest parasites on the other hand may be almost obscured by the prominent stippling of the corpuscle. They are called Schuffner's dots (Fig. 1) and they are found also accompanying monkey parasites of the tertian type

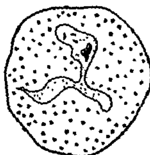


FIG. 1.



FIG 2.

(*P. cynomolgi, P. gonderi,* etc). *P. falciparum*—the parasite of malignant tertian malaria—causes another kind of stippling, called Maurer's clefts (Fig. 2). These are few in number and consist of bigger, irregular dots, circles and clefts. The immature

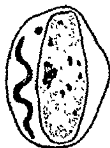


FIG. 3

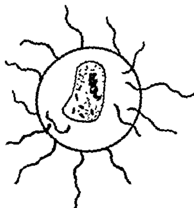


FIG 4.

sexual forms of this parasite are accompanied by very strange bodies in the erythrocytes, deeply staining loops, coils or bars—usually one only per infected cell (Fig. 3). I am working at present with Dr. R. B. Heisch on a malaria parasite of East African bats, which exhibits, instead of stippling, very prominent

filaments, emerging from all over the infected erythrocyte (Fig. 4). None of these phenomena is visible in the unstained preparation; all are easily demonstrated by heavy Romanowsky staining of fresh films—beyond these facts we know little about them, but they should not remain an insoluble problem.

Sporogony in the Mosquito

The mid-gut of a mosquito is the site of development of the male and female gametocytes we left last in the blood stream of the vertebrate. The male throws out about 6 or 8 flagella—gametes which lash about and eventually break free. They go in search of the female, which in the meantime has undergone some process of maturation, in which possibly polar bodies are extruded. But the exact nature of this nuclear change is unknown—in fact, at no step during the entire life cycle of the parasite is the nature of nuclear division understood. Does mitosis occur? No convincing evidence of the existence of chromosomes has been found. Even in the earliest days of malaria research this problem attracted attention; Bastianelli and Bignami (1899) thought that they were able to distinguish chromosomes in the course of the maturation of the microgametocyte. Chen (1944) and others have demonstrated by the Feulgen technique chromatin bands, which were assumed to be chromosomes. There are a few curious and possibly significant features about the nucleus at different stages which I shall mention later.

Fertilization of the female takes place in the contents of the mid-gut; the zygote a [illegible] the motile ookinete— [illegible] the surface of the mid- [illegible] motility? We don't know. Why is the nucleus of the zygote so extraordinarily dense, and why does it not begin to divide in the oocyst for a day or two? Eventually division takes place and a thousand or more sickle-shaped bodies—the sporozoites—form inside the oocyst. This ruptures and the sporozoites 'make their way' (but how?) to the salivary glands of the mosquito. When the mosquito bites the animal host, sporozoites enter with the saliva and a new cycle begins The cycle in the mosquito lasts from 7 days to a month or more.

How the microgametes, the ookinetes and the sporozoites move is a mystery: there are no muscle fibres, organs of locomotion or amoeboid movement.

Different species of mosquito vary in their susceptibility to infection. As far as we know, only mosquitoes of the genus *Anopheles* can carry human malaria. Even strains of the parasite vary in their behaviour in the insect; the tropical African variety of *P. falciparum* practically fails to develop in *A. maculipennis*—yet European strains of the parasite grow easily. The natural host of *P. berghei* is *A. dureni*—a mosquito confined to the Congo forests. In other species of *Anopheles* development of *P. berghei* usually stops short in the oocyst stage—possibly as the result of some food deficiency in these 'wrong' species of *Anopheles*. (Professor MacDonald has just drawn my attention to some recent American work—Terzian *et al.* (1952)—which is tending to the conclusion that the addition of para-aminobenzoic acid and other substances to the diet of the mosquito leads to greater [illegible] ment of malaria parasites in [illegible] my own [illegible] development of the parasite in I obtain[illegible] *A. maculip*[illegible] up to the oocyst stage—the oocysts became filled with sporozoites, ruptured, but the contents, instead of going to the salivary glands, just vanished into thin air—I assume they were lysed by some lethal substance in the haemocoelome.

Development of the Sporozoite in the Vertebrate

This is the tissue stage of the malaria parasite—or so-called exo-erythrocytic schizogony. The type of development of the sporozoite after it has entered the body depends upon the species of parasite. There are wide differences of behaviour amongst the different forms, though all end in the same way by invasion of the red blood corpuscles (Fig. 5).

In lizard malaria and in one form of bird malaria (*P. elongatum*) the sporozoite grows in various wandering cells and macrophages. the bone-marrow may be grossly infiltrated with exo-erythrocytic parasites, and the bird dies of a fatal anaemia.

In the common malaria of birds (*P. gallinaceum*, *P. relictum*), development takes place in the endothelial cells lining the

P. elongatum

P. gallinaceum

(P) kochi

P. vivax

♂

Fig. 5

capillaries. These cells become swollen and the vessels become blocked by large schizonts. This process is commonest in the brain and the bird dies of 'cerebral malaria'.

In the form of monkey malaria that I first studied—*Hepatocystis* (=*Plasmodium*) *kochi*—development of the parasite takes place in the parenchyma cells of the liver—a schizont grows and eventually a cystic structure containing millions of merozoites is produced. The cyst measures 2 or more mms. and is easily visible on the surface of the liver to the naked eye.

In human malaria to begin with the process is much the same as in *H. kochi*—a parenchyma cell of the liver is invaded by the sporozoite,[1] the growing parasite pushes aside the nucleus of the host cell, which increases in size to accommodate the schizont. The nucleus of the parasite divides many times, and at last (in 6–12 days according to species) a spherical or irregular body containing thousands of ripe merozoites is formed. *This body—the pre-erythrocytic schizont—bursts, the merozoites are discharged into the sinusoids of the liver and invade red blood corpuscles. This is the end of the incubation period.* In quartan and benign tertian malaria, a few merozoites return to parenchyma cells to initiate the exo-erythrocytic or relapse cycle. In malignant tertian malaria this does not apparently happen and there are no true relapses in this form of the disease.

The liver cycle in quartan and benign tertian apparently continues quite independently of what is happening in the blood, the degree of immunity, etc. Merozoites are discharged at intervals and when immunity is high, they are immediately destroyed; when immunity sinks, the merozoites are once again able to multiply in the red blood corpuscles and a relapse ensues.

As Fairley (1949) pointed out in his Croonian lectures, there are some puzzles still awaiting solution in this story. In some *cases of benign tertian malaria, the incubation period is prolonged for six months or so.* What is happening in the liver? No immunity could be present, so one might assume either that the sporozoite is arrested in its development or that only liver-invading merozoites are produced during this period.

[1] It is possible though unlikely that the sporozoite enters a reticulo-endothelial cell in the liver, giving rise to a primary generation of merozoites which then invade the parenchyma.

So far, tissue forms of only *P. falciparum* and *P. vivax* have been found experimentally in the liver of man. We are lucky in being able to work on the closely allied monkey parasites—*P. cynomolgi* which almost exactly resembles the human benign tertian parasite, and *P. inui* practically the same as the human quartan parasite. Details of exo-erythrocytic schizogony can be worked out with these simian parasites. Unfortunately there is no known counterpart to *P. ovale*, and the nature of its tissue forms will require the employment of a human volunteer.

I think it is worth while to give you a description of the methods we are now employing in working on this problem in mammalian malaria. The first thing to remember is the size of the liver and the dose of the inoculum. It is desirable to use baby monkeys (about 2 lb. in weight) in order to have a minimum of liver cells and the largest possible number of sporozoites. Then the pre-erythrocytic schizonts should easily be rendered visible. On the contrary, in nature, in an adult man, bitten by one or two mosquitoes, not particularly heavily infected, what chance is there of ever finding the forms? A biopsy or autopsy of such a man would only reveal the pre-erythrocytic schizonts, if at all, during the two or three days that they are large enough to be easily seen. How does one get the 'largest possible number of sporozoites'? One must have a colony of suitable anophelines of really large size—capable of giving at any given moment two or three thousand females. Good gametocyte carriers are the next essential and these have to be carefully sought for and used at the right moment. Repeated feeds over a week or so are desirable in order to get a 100 per cent infection rate in the insect. The mosquitoes are then kept at the appropriate temperature and humidity for two or three weeks (according to species), until the sporozoites are through to the glands. The best method of getting the sporozoites to the liver is to dissect out the glands in chilled Locke-serum mixture, grind them gently in a mortar and inoculate the suspension intravenously. Dissection of an adequate number of glands means the employment of four or more teams of trained workers, one team being able to deal with a maximum of 100 mosquitoes per hour.

Biopsies of liver are then made by open operation at appropriate

capillaries. These cells become swollen and the vessels become blocked by large schizonts. This process is commonest in the brain and the bird dies of 'cerebral malaria'.

In the form of monkey malaria that I first studied—*Hepatocystus* (=*Plasmodium*) *kochi*—development of the parasite takes place in the parenchyma cells of the liver—a schizont grows and eventually a cystic structure containing millions of merozoites is produced. The cyst measures 2 or more mms. and is easily visible on the surface of the liver to the naked eye.

In human malaria to begin with the process is much the same as in *H. kochi*—a parenchyma cell of the liver is invaded by the sporozoite,[1] the growing parasite pushes aside the nucleus of the host cell, which increases in size to accommodate the schizont. The nucleus of the parasite divides many times, and at last (in 6–12 days according to species) a spherical or irregular body containing thousands of ripe merozoites is formed. This body—the pre-erythrocytic schizont—bursts, the merozoites are discharged into the sinusoids of the liver and invade red blood corpuscles This is the end of the incubation period. In quartan and benign tertian malaria, a few merozoites return to parenchyma cells to initiate the exo-erythrocytic or relapse cycle. In malignant tertian malaria this does not apparently happen and there are no true relapses in this form of the disease.

The liver cycle in quartan and benign tertian apparently continues quite independently of what is happening in the blood, the degree of immunity, etc. Merozoites are discharged at intervals and when immunity is high, they are immediately destroyed; when immunity sinks, the merozoites are once again able to multiply in the red blood corpuscles and a relapse ensues.

As Fairley (1949) pointed out in his Croonian lectures, there are some puzzles still awaiting solution in this story. In some cases of benign tertian malaria, the incubation period is prolonged for six months or so. What is happening in the liver? No immunity could be present, so one might assume either that the sporozoite is arrested in its development or that only liver-invading merozoites are produced during this period.

[1] It is possible though unlikely that the sporozoite enters a reticulo-endothelial cell in the liver, giving rise to a primary generation of merozoites which then invade the parenchyma.

nothing else, that I have made it clear that the malaria parasite is endowed with a purpose and sense of direction that many of us might envy.

REFERENCES

[illegible]

CHEN, T. (1944). *Amer. J. Hyg.* 40, 26.

FAIRLEY, N. H. (1949). *Brit. med. J.* ii, 825.

GRASSI, B., BIGNAMI, A. and BASTIANELLI, G. (1898). *Atti R. Accad. Lincei. Rendic*, 5 ser., 8, i, 21.

HAWKING, F. (1951). *Brit. med. Bull.* 8, 16.

JAMES, S. P. (1931). *Trans. R. Soc. Trop Med. Hyg.* 24, 477.

LAVERAN, A. (1888). *Bull. Acad. Med* 9, 1235.

RAFFAELE, G. (1934). *Riv. di Malariol* 13, 332.

RAFFAELE, G. (1936). *Ibid* 15, 318.

ROMANOWSKY, D. (1891) *St. Petersburger Med. Woch.* 34, 297.

ROSS, R. (1897). *Brit med. J.* ii, 1786.

SHORTT, H. E. and COOPER, W. (1948) *Trans R. Soc. Trop. Med. Hyg.* 41, 427.

SHORTT, H. E and GARNHAM, P. C. C. (1948). *Ibid.* 41, 785.

SHORTT, H. E., FAIRLEY, N. H., COVELL, G., SHUTE, P. G. and GARNHAM, P. C. C. (1951). *Ibid* 44, 405.

SHORTT, H. E., COOPER, W and BRAY, R. S. (1953). *Ibid* 47, 11.

TERZIAN, L. A., STAHLER, N. and WARD, P. A. (1952). *J. Infect. Dis.* 90, 116.

intervals. So far, the earliest forms have been found at two days by Shortt, Cooper and Bray (1953). I have been doing an interesting extension of the technique lately in an attempt to discover the exact nature of the final process of schizogony. We had noticed in various species that the older schizonts seemed to be divided into distinct large segments before the final division into merozoites. We have called these cytomeres or pseudo-cytomeres. In the usual forms of malaria, growth of the parasite is so rapid that the process could not be properly watched. The quartan parasite however grows much more slowly and in my last experiments I have taken biopsies at 6-hourly intervals over a period of 48 hours. Each biopsy entailed a separate laparotomy, but the monkey stood the procedure quite well, with the aid of plasmosan. I have not finished examining the material, but have confirmed the presence of cytomeres, looking like cart-wheels with rims studded with nuclei. A curious feature is the change in the staining reaction of the nuclei—in the earlier and larger ones, they stain with relative difficulty with Romanowsky; the later ones, including the merozoite stage, stain so intensely that it is difficult to get good differentiation. It looks as though some sudden chemical change has occurred.

The last point of technique concerns the section. Fixation in Carnoy, followed by Colophonium-Giemsa staining, devised by Shortt and Cooper (1948), gives the prettiest results. Then comes the examination, and it may be that hundreds of sections will have to be searched, before a tissue form is seen. On the other hand, with some of our heavy inocula, we have encountered as many as nearly 50 schizonts per section.

I have been unable to describe details of the tissue stages of the different species, which, by the way, in human malaria, are confined to the parenchyma cells of the liver. Nor have I made any attempt to mention the new methods of [illegible] sio- [illegible] ost. logical requirements of the [illegible] rto Tissue cul[illegible] ma [illegible] on the other hand has easily [illegible] parasites in tissue culture.

In this quick review of the life history, I think, if I have done

standing. The average values for cardiac output in the recumbent position vary from 18 to 30 ccs. per min. per lb. of body weight or from 2 to 3 litres per sq. metre of body surface per minute, but there is great variation between the limits of equivalent body weights or values of body surface. In patients recovering from prolonged illness or operations the basal recumbent value may be low, i.e. 9 ccs./min./lb. of body weight.

On standing or on tilting upright, the normal subject may show a fall of as great as 2 L./min. in cardiac output; in general, the greater the recumbent value, the greater the fall, and indeed, where low normal recumbent values are recorded, there may be little initial change on tilting.

This fall in output is usually accompanied by an increase in oxygen consumption and in pulse rate; the concomitant reduction in stroke volume, accompanied by changes in the shape and size of the heart, denotes a greater oxygen utilization. Ventilation rate and volume are also increased.

If blood pressures be accurately measured either by strain gauge or capacitance manometry, it is found that mean pressures at aortic valve level do not normally change on standing. Some subjects however show a fall in mean pressure and others a

corresponds to the new difference in hydrostatic height. This fall in pressure presumably *initiates the carotid sinus reflex which* serves to maintain the mean blood pressure at heart level by vasoconstriction at the extremities and on the skin of the trunk and by tachycardia.

through the muscles of the leg, and is further responsible for the thick muscular coat of the vessels.

In the recumbent posture, Bazett and his co-workers (1935) *have shown that the rate of transmission of the pulse wave not* only increases from the central to the peripheral vessels but is greatly affected by assumption of the erect posture; there is a

XVIII

The Physiological Effects of Gravity

W. K. STEWART

To the clinician and physiologist, the term gravity is synonymous with the erect posture. Gravity is a force and the effects of gravity upon a body are due to the mass of the body or its component parts, and the acceleration of gravity (1 g=32 feet per sec. per sec.). Of recent years, the effects of gravity upon the circulation in particular in normal and abnormal subjects, have aroused a great deal of interest and also in situations where the value of the acceleration is a multiple of normal, as encountered in aircraft performing tight turns or loops.

In his *Monograph on Veins*, Franklin (1937) states that Lower clearly directed attention to the importance of gravity in 1669, citing the effect on large varicosities and in the causation of syncope. Modern concepts commence with the series of papers by Leonard Hill (1909) in the *Journal of Physiology*, published around the turn of the century, when the importance of the splanchnic area was clearly established in the body's compensatory reactions to the assumption of the erect posture.

Effects of Erect Posture

Between the two world wars, a great deal of literature became available. It is not the intention to review this in any detail but to summarize the major points of interest so as to obtain a general picture of the effects of gravity on the assumption of the erect posture.

Most normal subjects show a decrease in cardiac output on

Howard and his co-workers have found that the venous pressure at the ankle may rise to 130 mm. Hg., apparently sufficient to sustain a column of blood to head level of the subjects concerned. On other occasions in the same subjects, the hydrostatic venous pressure during quiet standing only supported the blood column to the level of the third interspace. The extra pressure appears to be derived from differences in muscular tension on the assumption of the posture. Further, in rocking backwards and forwards without moving the feet, the venous pressure fluctuates markedly with pressure differences of 45–50 mm. Hg. (60–70 cms. saline).

The time taken for the rise of venous pressure at the ankle depends largely on body and surface temperature but may be two minutes before it becomes stabilized in quiet standing. This is entirely due to rate of inflow into the veins and the action of the valves, since simultaneous recording of pressure above the valves, in the iliac vein, demonstrates a pressure rise which is synchronous with the increase in hydrostatic height on assumption of the erect, from the recumbent posture.

If the veins be previously congested by the application of a high intrathoracic pressure, the increase in pressure at the ankle is synchronous with the hydrostatic increase of pressure in the iliac vein.

The true function of the valves of the leg veins can therefore be seen to maintain blood flow during great variations of internal pressure, to assist in avoiding stasis in the capillaries of muscles and skin and to promote flow through the arterio-venous anastomoses.

The fluctuations of pressure concerned in muscular movement have been beautifully demonstrated by Pollack and Wood (1949) during walking. Here the musculo-respiratory pump is most active since the mean venous pressure at the ankle falls to 24 mm. Hg. from the average pressure of 86 6.

With continued marching, however, as in continued standing, this decrease in venous pressure is not sufficient to prevent an increase in foot volume which may amount to 100 ccs., and which does not decrease overnight (Whittingham, 1951), in contradistinction to Atzler and Herbert's (1923) findings during quiet standing.

slight reduction in the heart to subclavian region and a progressive increase to the periphery from 9 to 14 m./sec. in the femoral to dorsalis pedis arteries.

It is thought that these changes serve with the arteriolar constriction to diminish the effects of the erect posture on the capillary pressure in the foot. Although normal capillary pressure tends to be affected by the permeability of the walls it is also affected by the pressure in the veins as well as the loss of pressure head and flow in the arterioles. The pressure gradient in a capillary (from arteriole to venule) is dependent upon the rate of flow and the size of the lumen and, hence, where there is a rise in venous pressure, there is apt to be a decrease in flow and pressure gradient.

The effect of gravity on the venous system is well known. Measurement of the mean right auricular pressure shows a fall from a slight positive pressure to a slight negative pressure, with reference to ambient atmospheric pressure, on assuming the erect posture. Carrier and Rehberg (1923) thus found that in the dependent arm or leg, the venous pressures were below the calculated hydrostatic pressure. Pollack and Wood (1949) found that in the quiet standing position the average venous pressure in the ankle was 87 mm., the decrement from the calculated hydrostatic pressure being 2 mm. Hg. In their subjects, this pressure was sufficient to support a column of blood approximately to the level of the third interspace at the sternum.

The vasoconstriction which occurs on standing is not observed in a sympathectomized forearm. In postural hypotension, many studies of blood pressure, peripheral blood flow and cardiac output also demonstrate that the fundamental abnormality is loss of reflex vasoconstriction in response to the fall in carotid blood pressure. This results in a reduction in venous return to the heart, and in cardiac output of up to 40 per cent.

Brigden, Howarth and Sharpey-Schafer (1950) have shown that in some cases there may be a mechanical blockage to inferior vena cava flow at the level of the diaphragm, resulting in a marked rise in pressure below the heart to above hydrostatic values, with a consequent further fall in right auricular pressure and eventually fainting.

Howard and his co-workers have found that the venous pressure at the ankle may rise to 130 mm. Hg., apparently sufficient to sustain a column of blood to head level of the subjects concerned. On other occasions in the same subjects, the hydrostatic venous pressure during quiet standing only supported the blood column to the level of the third interspace. The extra pressure appears to be derived from differences in muscular tension on the assumption of the posture. Further, in rocking backwards and forwards without moving the feet, the venous pressure fluctuates markedly with pressure differences of 45–50 mm. Hg. (60–70 cms. saline).

The time taken for the rise of venous pressure at the ankle depends largely on body and surface temperature but may be two minutes before it becomes stabilized in quiet standing. This is entirely due to rate of inflow into the veins and the action of the valves, since simultaneous recording of pressure above the valves, in the iliac vein, demonstrates a pressure rise which is synchronous with the increase in hydrostatic height on assumption of the erect, from the recumbent posture.

If the veins be previously congested by the application of a high intrathoracic pressure, the increase in pressure at the ankle is synchronous with the hydrostatic increase of pressure in the iliac vein.

The true function of the valves of the leg veins can therefore be seen to maintain blood flow during great variations of internal pressure, to assist in avoiding stasis in the capillaries of muscles and skin and to promote flow through the arterio-venous anastomoses.

The fluctuations of pressure concerned in muscular movement have been beautifully demonstrated by Pollack and Wood (1949) during walking Here the musculo-respiratory pump is most active since the mean venous pressure at the ankle falls to 24 mm. Hg. from the average pressure of 86 6.

With continued marching, however, as in continued standing, this decrease in venous pressure is not sufficient to prevent an increase in foot volume which may amount to 100 ccs., and which does not decrease overnight (Whittingham, 1951), in contradistinction to Atzler and Herbert's (1923) findings during quiet standing.

slight reduction in the heart to subclavian region and a progressive increase to the periphery from 9 to 14 m./sec. in the femoral to dorsalis pedis arteries.

It is thought that these changes serve with the arteriolar constriction to diminish the effects of the erect posture on the capillary pressure in the foot. Although normal capillary pressure tends to be affected by the permeability of the walls it is also affected by the pressure in the veins as well as the loss of pressure head and flow in the arterioles. The pressure gradient in a capillary (from arteriole to venule) is dependent upon the rate of flow and the size of the lumen and, hence, where there is a rise in venous pressure, there is apt to be a decrease in flow and pressure gradient.

[illegible] reference to ambient atmospheric pressure, on assuming the erect posture. Carrier and Rehberg (1923) thus found that in the dependent arm or leg, the venous pressures were below the calculated hydrostatic pressure. Pollack and Wood (1949) found that in the quiet standing position the average venous pressure in the ankle was 87 mm., the decrement from the calculated hydrostatic pressure being 2 mm. Hg. In their subjects, this pressure was sufficient to support a column of blood approximately to the level of the third interspace at the sternum.

The vasoconstriction which occurs on standing is not ob-

[illegible] blood pressure. This results in a reduction in venous return to the heart, and in cardiac output of up to 40 per cent.

Brigden, Howarth and Sharpey-Schafer (1950) have shown that in some cases there may be a mechanical blockage to inferior vena cava flow at the level of the diaphragm, resulting in a marked rise in pressure below the heart to above hydrostatic values, with a consequent further fall in right auricular pressure and eventually fainting.

Howard and his co-workers have found that the venous pressure at the ankle may rise to 130 mm. Hg., apparently sufficient to sustain a column of blood to head level of the subjects concerned. On other occasions in the same subjects, the hydrostatic venous pressure during quiet standing only supported the blood column to the level of the third interspace. The extra pressure appears to be derived from differences in muscular tension on the assumption of the posture. Further, in rocking backwards and forwards without moving the feet, the venous pressure fluctuates markedly with pressure differences of 45–50 mm. Hg. (60–70 cms. saline).

The time taken for the rise of venous pressure at the ankle depends largely on body and surface temperature but may be two minutes before it becomes stabilized in quiet standing. This is entirely due to rate of inflow into the veins and the action of the valves, since simultaneous recording of pressure above the valves, in the iliac vein, demonstrates a pressure rise which is synchronous with the increase in hydrostatic height on assumption of the erect, from the recumbent posture.

If the veins be previously congested by the application of a high intrathoracic pressure, the increase in pressure at the ankle is synchronous with the hydrostatic increase of pressure in the iliac vein.

The true function of the valves of the leg veins can therefore be seen to maintain blood flow during great variations of internal pressure, to assist in avoiding stasis in the capillaries of muscles and skin and to promote flow through the arteriovenous anastomoses.

The fluctuations of pressure concerned in muscular movement have been beautifully demonstrated by Pollack and Wood (1949) during walking. Here the musculo-respiratory pump is most active since the mean venous pressure at the ankle falls to

[illegible]
increase in foot volume which may amount to 100 ccs., and which does not decrease overnight (Whittingham, 1951), in contradistinction to Atzler and Herbert's (1923) findings during quiet standing.

During normal life on the ground, the dramatic effects of gravity are therefore seen either in those cases where sympathetic reflexes fail or in quiet standing in the erect posture, where fainting may occur from mechanical constriction of inferior vena cava or by a general inadequacy of the circulation with *pooling of blood in distended veins and filtration of fluid from* capillaries, particularly in hot weather.

EFFECTS ON AIRCRAFT PILOTS

subjective effects upon the pilot of gross fatigue, mechanical impairment of muscular function, loss of vision, loss of consciousness, and impairment of psychomotor reactions.

The objective physiological effects are due to the relative increase in weight of the tissues of the body, which of course is proportional to the increase in acceleration. Where the radial acceleration in the turn is a multiple of gravity, i.e. 5 times 'g' or 5 g, the tissues of the face for example weigh five times as great as normal and not unnaturally sag (Stewart, 1945a). Investigations have been carried out both in flight and in man-carrying or animal centrifuges. There are limitations to both methods but the physiological data provided from centrifuges

flying personnel, is intrinsically similar to the amaurosis produced by external pressure upon the eyeball, and is apparently a progressive loss of contrast over the whole visual field. With a fixed visual field and maintenance of the acceleration at a constant level, the time taken for the visual field to disappear is proportional to the acceleration, the curve being a rectangular hyperbola (Craik and Stewart). Thus radial acceleration produces ischaemia of the retina and this ischaemia is due to a fall in carotid blood pressure, caused primarily by the increase in weight of the column of blood from the heart to the optic artery.

The eye is particularly susceptible to such a fall in pressure because of its intra-ocular pressure of 20–25 mm. Hg. In general terms, *the retinal ischaemia* has a *threshold* value of acceleration, provided that the duration of the acceleration be sufficienth [illegible] pressure in the caroti [illegible] lost if the [illegible]

Although subjects can be trained to show a reasonably constant threshold to acceleration, in practice a normally fit individual may show variations over ± 0·5 g. The fall in carotid pressure follows closely, in time, the onset of the acceleration and the loss of vision usually occurs some three to five seconds after the threshold value is reached. This time may therefore be related to the oxidative reserve of the retina. If however the value be markedly subthreshold, *no visual effects* may ensue although marked fatigue may be experienced. The reflex response initiated by the carotid sinus to the fall in carotid pressure may result in a rise in pressure some seconds after visual symptoms are experienced, and vision may tend to return; similarly if the acceleration be applied slowly, the resultant threshold is higher than if the peak acceleration had been reached within five seconds. The time for the carotid response would appear to be within 7–12 seconds in human subjects. Greenfield (1945) also observed a similar rise in carotid pressure in cats which was abolished by denervation of the sinus.

Unconsciousness is also produced at a value of acceleration about 0·7 to 1 g *higher than the threshold* value for *black-out*, but it is interesting to note that the average value of acceleration producing unconsciousness in R.A.F. flying personnel is higher than can be strictly accounted for by a fall in carotid blood pressure in a passive manner (Stewart, 1945b). There are two major reasons for this; firstly Lambert and Wood (1946) have shown that at the level of the third interspace the average decrease in arterial pressure was only 4 mm. Hg. for systolic pressure and 0 mm. Hg. for diastolic pressure. There was therefore no great effect upon the general peripheral resistance of the circulation. Secondly Martin and Henry (1951) have shown that there is a marked fall in jugular venous pressure during the

acceleration and hence a retention of a high arterio-venous differential, and it is only presumably when this is markedly reduced with a concomitant reduction in cerebral blood flow, that unconsciousness develops. On removal of the acceleration, the face which has been pale and blanched becomes suffused; the tachycardia is reversed and a marked bradycardia develops; the fall in arterial pressure is succeeded by a rise, frequently above normal control levels. Denervation of the carotid sinus reduces these latter effects.

The most interesting event is, however, the occurrence of convulsions, particularly seen in centrifuge studies, which are related to a preceding period of unconsciousness and to the stage of restoration of blood flow in the cerebral vascular tree and therefore largely to reoxygenation. Franks and his co-workers (1945) found that 52 per cent of 230 subjects had convulsions, when rendered unconscious in the R.C.A.F. centrifuge. Although the episodes were usually slight, severe generalized fits were also seen and were unrelated to abnormalities in resting electroencephalograms. There is still a great deal to be done in the investigation of these convulsions but bipolar leads did demonstrate the absence of spike and wave activity and no alteration of the unconscious pattern of slow high-voltage waves. Jasper and Cipriani (1945) however found in monkeys that during clinical episodes there was epileptiform activity in the E.E.G.

A great deal of investigation is still required therefore; the physiological changes which occur in other parts of the central nervous system are still unknown. If the episodes are truly subcortical in origin, the spinal cord may also be involved since, although it can be demonstrated that there are high negative pressures in the C.S F. within the cranium, there are most likely high positive pressures, relative to the atmospheric pressure, in the spinal theca. The circulation through the spinal cord under radial acceleration is consequently unknown and there also may be a great field of animal physiology to be explored here.

The reasons for the variations in threshold or resistance to the acceleration, within a young fit male population, still remain largely unexplored. For a time of exposure to the peak acceleration of 5 seconds, some individuals are found to lose

vision at 3 g and others at 9 g. In many cases, this variation in individual susceptibility may be found to be related to various active or passive methods of increasing the resistance. All these

Cinematographic films of the veins of the legs show marked congestion, but the total increase in volume is not sufficient to account for the loss of consciousness in the same terms as the syncope, which occurs in prolonged standing. The venous pressure in the veins of the leg shows a delayed response although it reaches the expected hydrostatic levels in animals with long durations of acceleration; above the valves the pressure increment is synchronous with the onset of the acceleration.

The most effective measures therefore are related to the promotion of a positive abdomino-thoracic differential pressure, thus promoting venous return to the thorax. Indeed the effectiveness of anti-g suits varies with the relative increase in intrarectal pressure which they produce. It has been found necessary to support the circulation in the legs and thighs as well as to raise the intra-abdominal pressure. The full physiological significance of this has still to be elucidated but it is undoubtedly related to flow of blood into the inferior vena cava. The importance of inadequate external pressure can also be demonstrated by the fact that if fluid pressure be used instead of air, very little protection is achieved unless the external water level reaches to the third interspace.

The scope of this lecture does not permit of discussion of many other facets of the physiology of gravity, as for example the head-down posture, or acceleration directed towards the head or the effect of gravity upon the otoliths or other systems of the body. Gravity is not only of great economic importance as it is manifested in normal life, but it will always represent a major challenge to the physiologist interested in aviation, who is already looking forward to the prospect of investigating the gravity-free state or the supergravity state that may be found in manned rocket flight.

acceleration and hence a retention of a high arterio-venous differential, and it is only presumably when this is markedly reduced with a concomitant reduction in cerebral blood flow, that unconsciousness develops. On removal of the acceleration, the face which has been pale and blanched becomes suffused; the tachycardia is reversed and a marked bradycardia develops; the fall in arterial pressure is succeeded by a rise, frequently above normal control levels. Denervation of the carotid sinus reduces these latter effects.

The most interesting event is, however, the occurrence of convulsions, particularly seen in centrifuge studies, which are related to a preceding period of unconsciousness and to the stage of restoration of blood flow in the cerebral vascular tree and therefore largely to reoxygenation. Franks and his co-workers (1945) found that 52 per cent of 230 subjects had convulsions, when rendered unconscious in the R.C.A.F. centrifuge. Although the episodes were usually slight, severe generalized fits were also seen and were unrelated to abnormalities in resting electroencephalograms. There is still a great deal to be done in the investigation of these convulsions but bipolar leads did demonstrate the absence of spike and wave activity and no alteration of the unconscious pattern of slow high-voltage waves. Jasper and Cipriani (1945) however found in monkeys that during clinical episodes there was epileptiform activity in the E.E.G.

A great deal of investigation is still required therefore; the physiological changes which occur in other parts of the central nervous system are still unknown. If the episodes are truly subcortical in origin, the spinal cord may also be involved since, although it can be demonstrated that there are high negative pressures in the C.S.F. within the cranium, there are most likely high positive pressures, relative to the atmospheric pressure, in the spinal theca. The circulation through the spinal cord under radial acceleration is consequently unknown and there also may be a great field of animal physiology to be explored here.

The reasons for the variations in threshold or resistance to the acceleration, within a young fit male population, still remain largely unexplored. For a time of exposure to the peak acceleration of 5 seconds, some individuals are found to lose

XIX

The Physiology of the Autonomic Nervous System

W. S. FELDBERG

It has taken me a long time to obtain a coherent and clear picture of the autonomic nervous system. Whether the picture is a correct one or not is a different matter. Some people, in fact, think it is not; but at least it has helped me in the past and I have not had reason to abandon it, and so I thought perhaps it might help others to look at this part of the nervous system in the way it appears to me.

Definition

The first task is to give a definition of the 'autonomic nervous system' because the term is not universally used in the same sense. I shall follow the definition of Langley which is the one generally used in this country. According to this definition, the autonomic nervous system consists of those fibres and nerve cells by means of which efferent impulses pass to the following histological structures. the gland cells, the smooth muscle and the heart. Thus the autonomic nervous system is the efferent peripheral pathway to all structures except skeletal muscle, or in other words, it is the efferent pathway to our viscera. This definition stresses two points: (*a*) the efferent character, and (*b*) the fact that we are dealing with a peripheral nervous system.

The efferent nerves can be sub-divided into somatic and autonomic. The somatic fibres to skeletal muscle pass as medullated nerve fibres from the anterior horn cell, in an uninterrupted pathway to the skeletal muscle (pathway No. 1, Fig. 1).

REFERENCES

ATZLER, E. and HERBERT, R. (1923). *Z. ges. exp. Med.* 38, 137.

BAZETT, H. C., COTTEN, F. S., LAPLACE, L. B. and SCOTT, J. C. (1935). *Amer. J. Physiol.* 113, 312.

BRIDGEN, W., HOWARTH, S. and SHARPEY-SCHAFER, E. P. (1950). *Clin. Sci.* 9, 79.

CARRIER, E. B. and REHBERG, P. B. (1923). *Skand. Arch. Physiol.* 44, 20.

CRAIK, K. J. W. and STEWART, W. K. (Unpublished observations).

FRANKLIN, K. J. (1937). *A Monograph on Veins.* Baillière, Tindall & Cox, London.

FRANKS, W. R., KEW, W. and ROSE, B. R. (1945). *J. Physiol.* 104 P.

GREENFIELD, A. D. M. (1945). *J. Physiol.* 104 P.

HILL, L. (1909). *Further Advances in Physiology.* Edward Arnold, London.

HOWARD, P., ROXBURGH, H. L. and STEWART, W. K. (Unpublished observations).

JASPER, H. H. and CIPRIANI, A. J. (1945). *J. Physiol.* 104 P.

LAMBERT, E. H. and WOOD, E. H. (1946). *Fed. Proc.* 5, 59.

MARTIN, E. E. and HENRY, J. P. (1951). *J. Av. Med.* 22, 382.

WHITTINGHAM, D. G. V. (1951). Thesis for Degree of M.D. Durham University.

XIX

The Physiology of the Autonomic Nervous System

W. S. FELDBERG

It has taken me a long time to obtain a coherent and clear picture of the autonomic nervous system. Whether the picture is a correct one or not is a different matter. Some people, in fact, think it is not; but at least it has helped me in the past and I have not had reason to abandon it, and so I thought perhaps it might help others to look at this part of the nervous system in the way it appears to me.

Definition

The first task is to give a definition of the 'autonomic nervous system' because the term is not universally used in the same sense. I shall follow the definition of Langley which is the one generally used in this country. According to this definition, the autonomic nervous system consists of those fibres and nerve cells by means of which efferent impulses pass to the following histological structures: the gland cells, the smooth muscle and the heart Thus the autonomic nervous system is the efferent peripheral pathway to all structures except skeletal muscle, or in other words, it is the efferent pathway to our viscera. This definition stresses two points: (*a*) the efferent character, and (*b*) the fact that we are dealing with a peripheral nervous system.

The efferent nerves can be sub-divided into somatic and autonomic. The somatic fibres to skeletal muscle pass as medullated nerve fibres from the anterior horn cell, in an uninterrupted pathway to the skeletal muscle (pathway No. 1, Fig. 1).

In contrast, the autonomic peripheral pathway is characterized by consisting not of one neurone but of two neurones; that is there is a ganglion or a cell station between the central nervous system and the periphery where one neurone ends and a new

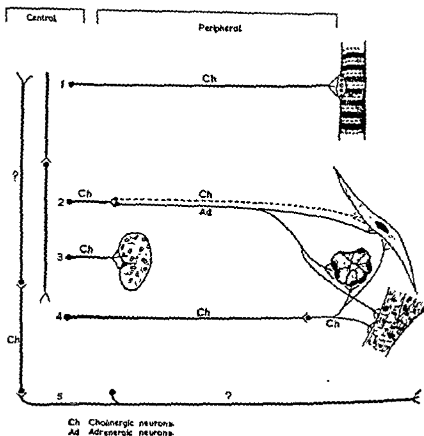


FIG 1. Diagram of the nervous system to illustrate the arrangement of cholinergic and adrenergic neurons (after Feldberg, 1950).

neurone starts (pathways No. 2 and 4, Fig. 1). This junction is called a synapse, and the transmission of the impulse from the nerve ending to the ganglion cell is referred to as synaptic transmission.

We have, further, the afferent sensory nerve fibres (pathway No. 5, Fig. 1). Sometimes those from the viscera are included in

the term autonomic nervous system, probably because a nervous system requires an afferent as well as an efferent pathway and reflex centres. It is actually not important which convention we use, as long as we are clear about the content of our terminology. We shall not include the afferent fibres from the viscera in the term autonomic nervous system, because they do not differ from ordinary sensory fibres from other parts of the body. But when we use the term autonomic nervous system in this restricted sense, we have to realize that all so-called autonomic nerves (vagus, splanchnici, or chorda tympani) are mixed nerves containing efferent autonomic and afferent sensory fibres. Another point which sometimes leads to confusion is

[illegible]
root ganglia of the spinal nerves and represent the trophic centres of the sensory fibres running in these nerves. No synaptic transmission occurs in these ganglia. In fact, it is not even certain that the nerve impulse passes through these ganglia which in some cases are situated in such a way that they can be cut out without severing the nerve. After such section, as long as the nerve fibres have not degenerated, a nerve impulse can be sent along the nerve and will cross the cut region.

Anatomical Considerations

The autonomic nervous system is an efferent pathway, but not all motor routes or cranial nerves which leave the central nervous system contain autonomic fibres. The outflow of autonomic fibres from the central nervous system is confined to certain regions. We distinguish a cranial, a sacral and a thoraco-lumbar outflow. There are some variations in different species as to how far the thoraco-lumbar outflow extends upwards and downwards, and the same applies to the sacral outflow. In these variations we are not interested at present. The sacral and cranial outflow or division is called the parasympathetic nervous system, and the thoraco-lumbar outflow the sympathetic. The sub-division of the autonomic nervous system into sympathetic and parasympathetic is thus purely an anatomical one, and this

terminology gives no information about the nature of the nerve fibres, whether they are cholinergic or adrenergic. This we must not forget.

Most, although not all, of our viscera are supplied by both kinds of nerves, which often act antagonistically. This antagonism is so striking that when a student learns the autonomic nervous system for the first time, he is so much impressed by this antagonism that he easily receives a wrong conception of the function of the autonomic nervous system. The first impression I remember I had of the autonomic nervous system, when a student, was that our viscera are, so to speak, led by two reins and that when we want to increase the activity we pull on the one or stimulate one, when we want to inhibit the activity we pull on the other. For instance, the sympathetic stimulates the heart, the vagus inhibits the heart. Now this idea about the two reins for the regulation of the activity of each viscus is misleading. We must take the view that, on the whole, our viscera are regulated by one rein only.

In order to obtain the right understanding of how the two divisions of the autonomic system work, we have to consider their anatomical arrangement. I mentioned that the peripheral autonomic pathway consists of two neurones; that the nerve impulse has to pass one synapse, but only one. The problem was where were these synapses? This problem was solved by Anderson's and Langley's nicotine experiments. If nicotine is painted on a nerve fibre in continuity nothing happens, but if it is applied to a ganglion where a nerve fibre ends and a new neurone starts, it excites the ganglion cells and we observe the effects in the periphery. But after a time the ganglion cells become paralysed and the nerve impulse can no longer be transmitted across the synapse. Therefore, by painting ganglia with large doses of nicotine, and then stimulating before and beyond the ganglion, we can find out whether the nerve fibre forms a synaptic junction in this given ganglion, or passes through it without synaptic interruption. With this method Anderson and Langley were able to map out the anatomical arrangement of the autonomic nervous system.

I always liked the following limerick which I think a Cambridge

student, who heard Langley's lectures in Cambridge, once published in a journal which has unfortunately ceased publication.

There was a young man of East Anglia
Whose nerves were a tangle of ganglia,
As was plain to be seen
He abhorred nicotine,
That makes ganglia Langlier and Langlier.

These ganglion cells function as distributing centres, that is, a single impulse reaching a ganglion may spread to up to twenty or more postganglionic fibres. Now the characteristic organization of the sympathetic system is that many of these distributing centres are near the central nervous system, mainly in the sympathetic chain; that is, the sympathetic system is clearly organized for diffuse distribution of the nerve impulses. The innervation of the suprarenal medulla provides a further illustration of the fact that the sympathetic nervous system is, so to speak, little interested in localizing the nerve impulse, as it leads in this instance to the release of a hormone into the blood stream.

The medullary cells of the suprarenals correspond to sympathetic ganglion cells. Both are derived from the same parental cell of the neural crest. We have only one neurone in the peripheral pathway to the suprarenals (pathway No. 3, Fig. 1); the second neurone has, so to speak, changed into an endocrine gland cell.

The characteristic of the parasympathetic ganglia is that they *lie peripherally, either close to the innervated tissue or even* inside the tissue; that is, the spread of the impulse is localized. This difference in the anatomical arrangement of the two divisions gives us the fundamental difference in the action of the two systems. It provides, as we shall see, a clearer understanding of the physiological function of the two divisions than the antagonism.

For instance, if we take the pupil: when light falls into our eye the pupil contracts; this is due to stimulation of parasympathetic fibres. If we enter a dark room the pupil dilates; this is not due to stimulation of the sympathetic but to inhibition of the parasympathetic. That is, if the parasympathetic is out of

action, the light reflex is absent, but the widely dilated pupil of a frightened person is a sympathetically stimulated pupil, and this sympathetic reaction is built into the pattern of a general sympathetic discharge.

The Intermediate Ganglia

Reference must be made to one other anatomical feature, the so-called intermediate ganglia, located in the rami communicantes. They were first described by Wrete in 1935 and later by Pick and Sheehan (1946), and are of practical importance in surgery of the sympathetic nervous system.

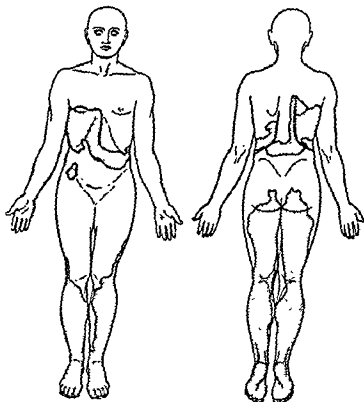
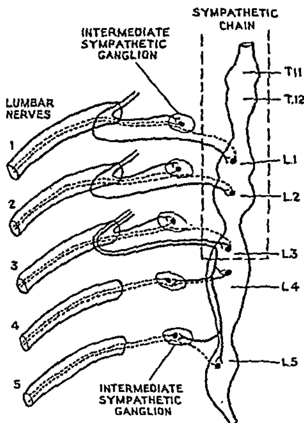


Fig. 2. Incomplete loss of thermoregulatory sweating in patient after thoracolumbar sympathectomy on both sides. Stippling shows areas of high electrical skin resistance indicating absence of sweating (after Boyd and Monro, 1949).

An operation of thoracolumbar sympathectomy, i.e. removal of the lower ganglia of the sympathetic chains, does *not* result in complete sympathetic denervation of the skin. The sympathetic



[illegible] glands and, if thoracolumbar sympa- [illegible]

[illegible] the body. This, however, is not the case, as shown [illegible] such a patient in a heating cabinet. In a normal person, this

procedure causes sweating over the whole body owing to stimulation of the sympathetic fibres to the sweat glands. After thoracolumbar sympathectomy, sweating is not completely absent from the lower part of the body; instead 'escape areas' or 'islands' of sweating constantly persist in the lower abdomen and front of the thighs, as shown in Fig. 2. In these regions the degree of sweating is less than before the operation, but its presence suggests that sympathetic denervation of the sweat glands is only partial.

Boyd and Monro (1949) have taken up this problem. They examined a large number of human embryos and found numerous sympathetic ganglia of varying size and distribution all along the rami communicantes, sometimes close to the lumbar nerves. These are the intermediate ganglia (Fig. 3). The position of many of these ganglia is such that they would necessarily escape removal in the usual operation of thoracolumbar sympathectomy. Boyd and Monro had the opportunity of confirming their results in an adult when examining the post mortem material from an operated patient who had shown such sudomotor escape areas. Intermediate ganglia are also found in the cervical region (Skoog, 1947) and may well be responsible for the atypical sudomotor responses seen after cervical sympathectomies.

Distribution of Cholinergic and Adrenergic Nerve Fibres

The nerve fibres of the autonomic nervous system act by the release of either acetylcholine or of noradrenaline and adrenaline. The first convincing experimental evidence in support of such chemical transmission of nerve effects was given by O. Loewi in 1920, in his now well-known classical experiments on the frog heart. He showed that stimulation of the parasympathetic fibres in the vagus to the heart released an inhibitory substance, now recognized as acetylcholine, and stimulation of the sympathetic accelerans fibres, a substance which, like adrenaline, stimulated the heart and is probably adrenaline itself. The nerve effects on the heart were due to an action of these released chemical transmitter substances.

Later work has shown that chemical transmission of nerve

effects is not restricted to the autonomic nervous system and that even here the anatomical designation of a fibre as belonging to the sympathetic or parasympathetic nervous system does not tell us whether it acts by the release of acetylcholine or of noradrenaline and adrenaline. A new terminology became necessary, designating a nerve fibre according to the transmitter substance it releases. This terminology was introduced by Dale (1933), and has been generally accepted. Since then we speak of *cholinergic* nerve fibres as neurones which act by the release of acetylcholine and of *adrenergic* nerve fibres as neurones which act by the release of noradrenaline and/or of adrenaline.

Research in the last years before the war was mainly concerned with the distribution of cholinergic and adrenergic neurones. In order to understand this distribution, we must keep in mind the fact that the peripheral pathway of the autonomic nervous system, with one exception, consists of two neurones. The results, which are illustrated in Fig. 1, can be summarized as follows.

The majority of the postganglionic, sympathetic fibres are adrenergic. This explains the similarity of the effects of noradrenaline and adrenaline with those of sympathetic nerve stimulation.

The postganglionic, parasympathetic fibres are cholinergic. This explains the similarity between the effects of acetylcholine and of parasympathetic nerve stimulation.

A few postganglionic, sympathetic fibres are cholinergic. It has long been known that the sweat glands in humans are supplied by fibres which anatomically belong to the sympathetic nervous system, but that they are activated by parasympathomimetic substances and that sympathetic sweat secretion is abolished by atropine. This anomaly is now cleared up; the postganglionic sympathetic fibres to the sweat glands of man, and the same applies to cats, are cholinergic. There are also cholinergic, sympathetic postganglionic vasodilator fibres to the skeletal muscles of dog and cat. In Fig. 1, these fibres are indicated by the discontinuous line.

There is, so far, only one known instance of postganglionic, parasympathetic fibres which are apparently adrenergic: the

procedure causes sweating over the whole body owing to stimulation of the sympathetic fibres to the sweat glands. After thoracolumbar sympathectomy, sweating is not completely absent from the lower part of the body; instead 'escape areas' or 'islands' of sweating constantly persist in the lower abdomen and front of the thighs, as shown in Fig. 2. In these regions the degree of sweating is less than before the operation, but its presence suggests that sympathetic denervation of the sweat glands is only partial.

Boyd and Monro (1949) have taken up this problem. They examined a large number of human embryos and found numerous sympathetic ganglia of varying size and distribution all along the rami communicantes, sometimes close to the lumbar nerves. These are the intermediate ganglia (Fig. 3). The position of many of these ganglia is such that they would necessarily escape removal in the usual operation of thoracolumbar sympathectomy. Boyd and Monro had the opportunity of confirming their results in an adult when examining the post mortem material from an operated patient who had shown such sudomotor escape areas. Intermediate ganglia are also found in the cervical region (Skoog, 1947) and may well be responsible for the atypical sudomotor responses seen after cervical sympathectomies.

Distribution of Cholinergic and Adrenergic Nerve Fibres

The nerve fibres of the autonomic nervous system act by the release of either acetylcholine or of noradrenaline and adrenaline. The first convincing experimental evidence in support of such chemical transmission of nerve effects was given by O. Loewi in 1920, in his now well-known classical experiments on the frog heart. He showed that stimulation of the parasympathetic fibres in the vagus to the heart released an inhibitory substance, now recognized as acetylcholine, and stimulation of the sympathetic accelerans fibres, a substance which, like adrenaline, stimulated the heart and is probably adrenaline itself. The nerve effects on the heart were due to an action of these released chemical transmitter substances

Later work has shown that chemical transmission of nerve

system. It has, however, long been known that the following two structures have much in common pharmacologically: the ganglion cells of the autonomic nervous system and the motor endplates of skeletal muscle. This pharmacological similarity is also shown in the innervation; the motor fibres to skeletal muscle are cholinergic.

The diagram of Fig. 1 illustrates that all efferent fibres which leave the central nervous system are cholinergic. We may further say that all peripheral efferent fibres are either cholinergic or adrenergic. The position is different with regard to the afferent fibres. They are neither cholinergic nor adrenergic. We do not know the transmitter for the sensory nerve impulses in the posterior roots to the nerve cells of the spinal cord. That brings us to the problem of the transmission in the central nervous system. Acetylcholine is not the transmitter of the first ascending sensory neurone, but there is good evidence that it exerts this function on certain central synapses. It is, however, not the universal central synaptic transmitter. Recent work suggests that the central nervous system is built up of cholinergic and non-cholinergic neurones which, in many but by no means in all instances, alternate with each other. This alternation is indicated in the diagram of Fig. 1.

The Function of the Parasympathetic Nervous System

If we look upon the parasympathetic system as the main efferent reflex pathway for viscera which are *localized* and not widely distributed over the whole body, we obtain a good understanding of the functioning of this division. Its fibres are the efferent pathway for the light and accommodation reflex. Cutting the parasympathetic fibres to the lachrymal glands abolishes reflex lachrymation. When we eat, or when we see food or think of it and our mouth waters, the salivary secretion results from stimulation of the parasympathetic nerves. But when a cat spits in rage it probably activates its sympathetic fibres to the salivary glands, the secretion presenting, then, one of the many features of the general emotional reaction pattern. As far as our gastro-intestinal secretion is under reflex control, the parasympathetic division provides the efferent pathway for the conditioned and

parasympathetic innervation to the cardia of the rabbit's stomach (von Brücke and Stern, 1938).

The next advance in our knowledge of the distribution of the cholinergic nerve fibres is best understood from the pharmacology of acetylcholine. This substance acts on gland cells, smooth muscles and the heart. These actions, which are abolished by small doses of atropine, are usually referred to as the muscarinic actions of acetylcholine, because muscarine, the poison of the common toadstool, exerts all these actions.

TABLE 1. Peripheral structures on which acetylcholine acts

(Muscarinic effects)	(Nicotinic effects)
gland cells	cells of suprarenal medulla
smooth muscles	autonomic ganglion cells
heart	motor end-plates of skeletal muscles

The structures affected by the muscarinic effects of acetylcholine are those innervated by the cholinergic, postganglionic, parasympathetic and sympathetic fibres. But after atropine, more atropine-resistant actions of acetylcholine are unmasked which resemble the actions of another drug, nicotine, and which Dale (1914) termed the nicotinic actions of acetylcholine. The structures affected are given in Table 1 and give us a clue where to look in our diagram, Fig. 1, for further physiological actions of acetylcholine when released by nerve endings. The synaptic transmission across a sympathetic ganglion is effected by acetylcholine. We can say that all preganglionic, sympathetic, and parasympathetic fibres are cholinergic

The embryological cells of the suprarenal medulla originate from the same parental cells as the cells of the sympathetic ganglia It is, therefore, not surprising that the nature of the innervating nerve fibres to the sympathetic ganglia and to the suprarenal medulla is the same. In other words, the sympathetic fibres to the suprarenal medulla are cholinergic. In fact, the cholinergic nature of these fibres was established before that of the other preganglionic, sympathetic and parasympathetic fibres.

So far we have confined ourselves to the autonomic nervous

rate per minute is less than 70 but increases to over 250 after section of the vagi. On the other hand, the pulse rate of the tame rabbit is about 200 but only increases to a little over 300 after section of the vagi. Racehorses, like athletes, usually have slow heart rates. The significance of this vagal brake is best understood when we realize that slowing of the heart, which is mainly accounted for by a lengthening of the diastole, means that a given amount of work is performed by the heart with less oxygen consumption, i.e. with less energy liberation, when it beats slowly than when it beats quickly.

The vagal tone of the heart is influenced continuously by many reflexes; for instance, when the heart is suddenly called upon to work harder, i.e. when the venous return increases or the arterial blood pressure rises. With increased venous return, the great veins and the right auricle become distended, and the rise in venous pressure elicits reflexly a quickening of the heart—the Bainbridge reflex. It is brought about by decrease of vagal tone; sympathetic stimulation contributes to a minor extent, if at all, to this acceleration. When the work of the heart is increased by raising the arterial resistance, the heart rate is slowed. This is actually the old observation which Marey (1863) made—often termed the 'Marey law'—that the heart rate varies inversely to the arterial pressure. We now know that it is brought about by reflexes elicited on stretching the aortic arch and the carotid sinus, the aortic arch and the carotid sinus reflexes, and resulting in increased vagal tone to the heart. In other words, the parasympathetic is again the efferent pathway.

Reflexes generally serve a useful purpose and the reflex slowing of the heart in response to increased arterial pressure is no exception to this rule. Its 'purpose' is best understood from the following findings: (1) A heart works more efficiently at a slow than at a quick rate. For instance, if 5 l. of blood per minute were ejected against a given peripheral resistance, the oxygen consumption would be smaller if the ejection were done with 50 instead of with 100 beats. (2) Increasing the work of the heart by raising the arterial pressure, but keeping the rate consistent, causes an increase in oxygen consumption which is nearly proportional to the increase in work (Evans and Matsuoka, 1915;

unconditioned reflexes. Concerning the autonomic innervation of the urinary bladder, the student often first learns that the parasympathetic contracts the detrusor muscle and relaxes the internal sphincter, whereas the sympathetic exerts the opposite effects. From this antagonism he endeavours to obtain an understanding of the reflex control of micturition. The bladder, however, is a localized viscus, and accordingly we find the parasympathetic as the efferent pathway for the different micturition reflexes, so far as they are not affected through the pudendal nerves. Injury of the parasympathetic leads to paresis of the bladder, whereas section of the sympathetic does not interfere with micturition. According to Barrington's analysis of the bladder reflexes, the sympathetic is only involved in the weak and transient reflex contraction of the detrusor muscle when the proximal part of the urethra is distended. Probably the main physiological function of the sympathetic innervation to the bladder is concerned with vasomotor reactions and with sex orgasm.

The innervation of the heart provides another instance in which the antagonism of the sympathetic and the parasympathetic fails to shed light on the physiological function of the two divisions of the autonomic nervous system. The statement that the sympathetic stimulates and accelerates, the parasympathetic inhibits and slows the heart gives too easily the impression that two reins regulate the heart action, whereas in fact reflexes which project on the heart as such, work mainly, if not wholly, through the parasympathetic. Through its fibres, which run in the vagus nerve, a continuous discharge is exerted upon the heart. This is shown by the fact that the heart rate in man may double when the 'vagal brake' is suddenly removed, as after atropine. At birth vagal tone is weak, in a new-born baby, for instance, atropine causes little increase in pulse rate. The degree of vagal tone is dependent on the way of living. The slow heart rate of the athlete, the so-called sinus bradycardia, is well known, and it is also known that schoolboys who train for endurance tests develop a slow pulse rate. Corresponding observations are made in animals adapted to endure prolonged exercise (Clark, 1927). The hare has a strong vagal tone; its pulse

defence and to produce those changes necessary for preparing the organism for 'fight and flight'. It is in this connection that Cannon referred to the emergency function of the sympatho-adrenal system, and it is also in the light of this protective mechanism that the effects of sympathetic stimulation are best understood and remembered.

Cannon has raised the interesting problem that the mechanism of sympathetic discharge may actually be harmful unless the emotion is transformed into action. Heart and circulation may be worked just as hard from an armchair as from a rower's seat. 'If no action succeeds the excitement, and the emotional stress—even worry and anxiety—persists, then the bodily changes due to the stress are not a preparatory safeguard but may be in themselves profoundly upsetting the organism as a whole.' Thus, the manner in which our sympathetic system reacts demands that we either take an objective attitude or permit our excite-

a child have a run in the garden when it gets excited may thus have a physiological meaning: that of 'working-off' the sympathetic discharge.

That a sympathetic discharge may be harmful in pathological conditions is well known clinically. In traumatic shock or in shock after haemorrhage, the sympathetic discharge, by constricting the vessels and impeding the flow through the tissues, may be so harmful as to lead to death. We have, therefore, to eliminate the reflex vasoconstriction which ordinarily sustains the blood pressure in order to ensure an adequate blood flow. Therapeutically we proceed by blocking the sympathetic ganglia with ganglion blocking drugs such as hexamethonium, or by preventing the adrenaline and noradrenaline released at the periphery from contracting the muscles of the vessel wall by the so-called anti-adrenaline substances, also called adrenergic blocking substances, such as dibenamine (for references see Nickerson, 1949).

The possibility of being able to activate the whole sympatho-adrenal system suggests a central representation of the

Evans, 1939). In other words, a rise in arterial blood pressure without slowing puts a great demand or load on the heart. But, owing to the increased vagal tone, the increase of work due to a rise of arterial blood pressure is often effected without any increase of oxygen usage at all. (Gollwitzer-Meier, Krotz and Krüger, 1938). Thus the heart works much more efficiently at the higher load than at the lower. The reflex slowing has therefore aptly been termed by Lovatt Evans 'the unloading reflex'.

In the Bainbridge as well as in the unloading reflex, the heart is influenced as an organ by itself and not as part of the general circulation. In all such heart reflexes the parasympathetic provides the efferent peripheral pathway. But reflex changes brought about in the heart in response to general circulatory demands, for instance in response to strenuous muscular exercise, involve the sympathetic as well.

The Function of the Sympathetic Nervous System

Its function is readily understood when we realize (1) that it can be activated as a whole—the so-called emergency function of the sympatho-adrenal system, but (2) that it is also the efferent reflex pathway for viscera which are widely distributed in the body, and (3) that tonic activity is a typical feature for a large fraction of the sympathetic system.

Activation of the whole sympathetic system. According to Cannon (1929) the sympathetic innervation fulfils an important function in making an animal fit for states of emergency. A widespread discharge of impulses occurs in the sympathetic system in the following conditions: during severe muscular work, in situations of danger, in extreme temperatures, asphyxia, haemorrhage, under strong emotions such as fear or rage, or when in pain. The discharge also affects the fibres to the suprarenal medulla, leading to an output of adrenaline and noradrenaline. We may look upon this widespread discharge as a kind of reflex action of the organism in response to a condition of stress. As mentioned before, the characteristic feature of many reflexes, like coughing or swallowing, is that they serve a useful purpose, the 'purpose' of the sympathetic discharge being to strengthen the powers of

are under sympathetic control. We may add another innervation, that to the liver whereby the blood sugar level can be controlled. As far as blood sugar changes are brought about by nervous mechanisms without interaction of hormonal factors (insulin, pituitary hormone and adrenaline), the sympathetic nervous system provides the pathway. Although the viscus responsible for the blood sugar changes is not widely distributed in the body, the result of the sympathetic stimulation to the liver, the hyperglycaemia, affects the whole organism.

The sympathetic forms the whole connection between the temperature regulating centre and the sweat glands and pilomotors. Heat loss is almost entirely dependent on the sympathetic nervous system, whether we consider the increased heat loss in man by radiation through vasodilatation or by evaporation of sweat, or whether we consider the regulation in animals by erection of the feathers or hairs.

Furthermore, it cannot be emphasized strongly enough that the sympathetic forms the sole connection or efferent pathway between the vasomotor centre and the blood vessels. How, then, is vasodilatation brought about? Dale once contrasted the main mechanisms of vasoconstriction and vasodilatation in the body as follows. vasoconstriction is usually a widespread effect centrally induced, vasodilatation a localized effect brought about peripherally, for instance, by vasodilator metabolites produced in the active tissues. This contrast, although it does not give the full picture of vasomotor control, stresses a fundamental difference in the mechanisms by which active vasoconstriction and vasodilatation are brought about in the body. Nevertheless, centrally induced vasodilatation is a frequent occurrence in the regulation of vasomotor tone, but brought about by inhibition of vasoconstrictor tone Any regulation of vasomotor tone through the vasomotor centre is via the sympathetic and is produced by changing the degree of the vasoconstrictor discharge. The parasympathetic does not participate in this regulation.

It is true that the parasympathetic contains vasodilator fibres to certain organs, such as the salivary glands, but these fibres do not take part in the regulation of vasomotor tone and cannot be activated from the vasomotor centre (Bernthal, Motley win

whole system. Such a centre is located in the hypothalamus and its electrical stimulation produces widespread sympathetic discharge, as first demonstrated by Karplus and Kreidl and later by Ranson and Magoun (1939). We have further representations in the cerebral cortex, and, on the other hand, also on lower levels in the medulla and spinal cord.

The fact that the system can be activated as a whole must not lead to the idea that it is always activated in this way. For instance, if we consider the sympathetic reaction pattern to increased environmental temperature, we find activation of the secretory fibres to the sweat glands with inhibition of the vasoconstrictors. On the other hand, the 'cold sweat' in strong emotions results from activation of both secretory and vasoconstrictor fibres to the skin. A typical instance of a limited sympathetic control is the emotional blushing which shows that the vasoconstrictor tone to one skin area can be inhibited without affecting the tone in another. In women who blush frequently and vividly, the blush area is usually confined to the face and to an area in the neck and sometimes the arms' areas of skin not covered by the cut of modern dress. Inhibition of vasoconstrictor tone thus occurs in these regions only. There is a discussion between Charles Darwin and Burgess on how this is brought about Burgess believed that it is due to specific anatomical arrangement of the facial vessels, whereas Darwin believed that the blush is confined to the face because the face is subjected to much closer and more earnest self-attention than any other part of the body. Darwin's argument revolved largely around mental concentration upon certain parts of the skin and he remarks that in races that wear little clothing and whose faces fail to attract undivided attention, blushing is less confined to these areas In other words, according to Darwin, those people who wear no clothes at all either do not blush or, if they do, they blush over the whole body, and this appears to be the case.

The sympathetic as efferent pathway of reflexes to viscera widely distributed in the body. Such viscera are the pilomotors, the sweat glands and the vessels. Body temperature and vasomotor tone

and Mendelhall, 1915; Freeman and Rosenblueth, 1931; Bacq, Brouha and Heymans, 1932, 1933, 1934; Bishop, Heinbecker and O'Leary, 1933; Tournade and Malméjac, 1933; Rosenblueth and Cannon, 1934; Pinkston, Partington and Rosenblueth, 1936; Back, 1945/46), but in most of the recent work no evidence could be obtained for Bayliss's idea of reciprocal vasomotor innervation (Jarisch, 1925; Bacq, Brouha and Heymans,

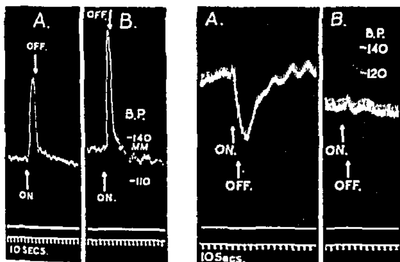


FIG. 4. Arterial blood pressure of normal (A and B on the left) and sympathectomized (A and B on the right) cat. Effect of central stimulation and of cut tibial nerve (A) before and (B) during curarine (3 mg. intravenously) (after Brown and Maycock, 1940).

1933; *Thomas and Brooks, 1935; Wybauw, 1938; Brown and* Maycock, 1940; Dole and Morison, 1940; Bernthal, Motley, Schwind and Weeks, 1941; Folkow, Strom and Uvnas, 1950; Celander and Folkow, 1951).

The experiments of Brown and Maycock are particularly striking because they illustrate, in addition, how deceptive *results can be. In the decerebrate cat, stimulation of the central* end of a sensory nerve, or brief occlusion of the vertebral arteries, produces vasoconstriction, whereas in the sympathectomized cat the response is mainly vasodilatation. In Fig. 4 is illustrated

and Weeks, 1945; Celander and Folkow, 1951). These vasodilator fibres provide a good example to show that the antagonism in the autonomic nervous system does not help us to understand the functions of the two divisions. Often the student first becomes acquainted with the autonomic innervation of the vessels when he learns of the two classical discoveries of Claude Bernard, who showed that the sympathetic contains vasoconstrictor fibres to the external ear of the rabbit, and that the parasympathetic contains vasodilator fibres in the chorda tympani to the salivary glands. From there it may take him a long time before he finds out that this simple antagonism of the two divisions does not provide him with a key for understanding vasomotor tone, and that the increase and decrease of tone are not the result of increased sympathetic and parasympathetic discharge respectively; in other words, no reciprocal vasomotor control exists. The parasympathetic vasodilators serve specific functions in strictly localized organs. For instance, without a sufficient supply of fluid, the salivary glands would be unable to keep up secretion; it is therefore not surprising that the pattern of the salivary reflex incorporates a localized vasodilatation, which is mediated by the parasympathetic. The sympathetic vasoconstrictor and the parasympathetic vasodilators are thus not antagonists for the same function but form efferent pathways for different reflex centres.

The Problem of Reciprocal Vasomotor Control

The opposite view of reciprocal vasomotor control has been put forward by Bayliss (1902, 1908). His authority on this subject was so great that his view is still widely held, although no longer supported by the bulk of the more recent experimental evidence. According to Bayliss, the sympathetic is not the sole efferent pathway from the vasomotor centre but reflex vasodilatation is brought about by activation of vasodilator fibres which belong to the parasympathetic as well as to the posterior root system.

The literature on this subject is rather confusing. There are a number of publications, none of them convincing, in support of this view (Mislavsky and Fofanow, 1908; Asher, 1909; Fofanow and Tschalussow, 1913, Tschalussow, 1913; Martin

response to released acetylcholine, and that inhibitory nerves to the heart in the original sense did not exist. The theory of chemical transmission of nerve effects has had the same impact on our approach to the problem of vasodilator nerves. The emphasis has shifted from the neurological to the pharmacological side, with the result that the term 'vasodilator nerves' has lost some of its clear and well-defined meaning, as seen from the following conclusions.

(1) A substance may have a vasodilator effect at one part of the vascular tree, say the capillaries, and a constrictor effect on another part, say the arterioles. If such a substance were released at these sites from different twigs of the same nerve fibre, this fibre would be vasodilator as regard to one twig and vasoconstrictor as regard to the other twig, but neither the term vasodilator nor vasoconstrictor nerve fibre would be applicable.

(2) A substance may cause vasodilatation when applied in a weak but vasoconstriction when applied in a somewhat stronger concentration. For our considerations, it would be irrelevant whether the effects were at the same or at different sites of the vascular tree, but if such a substance were released from nerve endings, the same nerve fibre could be considered to be vasodilator or vasoconstrictor according to the frequency of excitation.

(3) A substance may have different effects according to the condition of the vessel. Certain drugs may abolish or reverse the effects of another drug. If a nerve fibre acts by the release of a substance, this substance may therefore be vasoconstrictor or vasodilator according to the special state of the vessel at the moment of nerve stimulation. Again, it may be difficult to apply the term vasodilator or vasoconstrictor nerve fibre to such a special case. In the elucidation of vasodilator nerves, the reversal of vasoconstrictor effects by drugs has played an important role. The best known instance is Dale's reversal of the adrenaline response by ergotoxin (Dale, 1913). Such a reversal may mean either unmasking of vasodilator fibres or simply reversal of the action of a transmitter substance. Without knowing its nature and its behaviour under the action of the reversing drug, no safe conclusions are permissible.

the rise in arterial blood pressure in the normal and the depressor response in the sympathectomized cat on stimulation of the central end of a sensory nerve. The vasodilatation in the sympathectomized cat is not reflex in origin but can be accounted for by the muscular contractions which occur during stimulation. When these are prevented by curarine, stimulation of the central end of a sensory nerve or occlusion of the vertebral arteries no longer affects the blood pressure in the sympathectomized cat, as is illustrated in the figure.

In a few completely sympathectomized cats, Brown and Maycock found that transection of the spinal cord or ischaemia of the medullary centres caused a fall in arterial blood pressure, although the vagi had been cut and the animal fully curarized. They concluded that the effect might be conveyed to the periphery by some extra-sympathetic path, but the presence of intermediate ganglia would fully account for the result.

Muscular contractions and the presence of intermediate ganglia are two main sources of error which may simulate the existence of a vasomotor control through vasodilators outside the sympathetic nervous system.

Sympathetic Vasodilator Nerves to Skin and Skeletal Muscles

Our view that the vasomotor control is exerted solely through the sympathetic and that its fibres form the only efferent connection with the vasomotor centre does not exclude the possibility of vasodilator fibres within the sympathetic itself. Such fibres do indeed exist, and apart from the vasodilators to the coronary arteries which need not be considered in this connection There is good evidence of a sympathetic vasodilator supply to the limb muscles of various animals and man, and, in man, apparently also to the skin.

Before discussing these nerve fibres, it is worth while to examine the meaning of the term 'vasodilator nerves'. The problem is reminiscent of that of 'inhibitory nerves' to the heart which, in former times, led to interesting philosophical speculations. These lost their meaning when it was found that the inhibition was not a direct action of the nerve impulse but a

extremities. In these patients, cold induces and maintains spasm of the digital vessels. This spasm could be released by warming the body. For this purpose the patient was seated in a warming chamber, as illustrated in Fig. 5. A purse string was fitted round

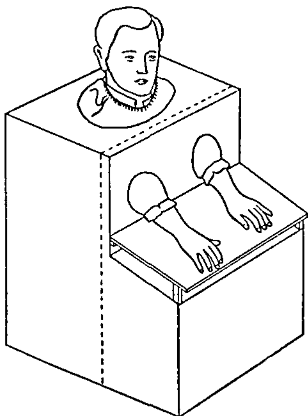


FIG 5. Warming chamber to record skin temperature of the hand and fingers whilst warming the body (after Lewis and Pickering, 1932).

his neck, his hands were resting on a cork plate and the skin temperature of his hand and fingers was recorded thermo-electrically. Warming the body raised the skin temperature in these regions This did not occur in patients after sympathectomy; therefore, the release of the vascular spasm must be transmitted via the sympathetic, but it could not have resulted from

The difficulty in defining vasodilator nerves would not exist if all sympathetic vasodilators were cholinergic. In fact, this is the conclusion to which Uvnas and Folkow in Sweden have come as the result of their extensive work on sympathetic vasodilators in animals. They assume that even the sympathetic vasodilators to the coronary arteries are cholinergic (Folkow, Frost, Haeger and Uvnás, 1948). Their view is opposed by Bulbring and Burn. Why these two schools came to different conclusions is understandable to a certain extent when we remember that in the period between Bülbring and Burn's experiments and those of the Swedish workers, our view about the nature of the transmitter of the adrenergic sympathetic vasoconstrictors has changed. Bulbring and Burn thought the transmitter was adrenaline; Folkow and Uvnas performed their experiments from the viewpoint that the transmitter is not adrenaline but noradrenaline which, unlike adrenaline, has scarcely any vasodilator component. What we can definitely say is, the sympathetic vasodilators to the limb muscles in dogs as well as in cats are cholinergic, but the transmitter for all other sympathetic vasodilators is not proved: it may be acetylcholine, it may not.

The question is, what is the physiological function of these sympathetic vasodilators to skeletal muscles and skin? Have we perhaps to modify our idea that central regulation of vasomotor tone is solely brought about by variation in sympathetic vasoconstrictor tone, and, in spite of what I said before, to assume that it is under the influence of two reins, but both belonging to the same system, the sympathetic? The evidence available suggests that this is not so, but that the sympathetic vasodilators are activated in those conditions which have been designated by Cannon as states of emergency, but that they do not participate in the reflex regulation of vasomotor tone through the vasomotor centre. This regulation is by one rein only, the sympathetic vasoconstrictors.

Sympathetic vasodilators to the human skin. The first evidence for the existence of such fibres was obtained by Lewis and Pickering (1932) on patients with Raynaud's disease in the upper

and Holling think it is more likely that the effect is brought about through nerve fibres supplied to the vessels themselves, that is, through activation of vasodilator fibres. If that is so, it is interesting to emphasize that this activation occurs only when the warming of the body is intense, in other words, in a condition of stress or in a state of emergency. A similar conclusion will be reached when we consider the sympathetic vasodilators to the skeletal muscles in man.

Sympathetic vasodilators to the skeletal muscles of man. Barcroft and Edholm (1945) showed that a sympathetic discharge into the limbs may result in cutaneous vasoconstriction, together with vasodilatation in the muscles. They studied the rate of blood flow into the forearm and hand as follows. The forearm or the hand is in a plethysmograph and a rubber armlet placed proximal to it. If pressure is thrown abruptly into the armlet, sufficient to impede the blood flow in the veins but insufficient to interfere with that in the arteries, the rate of increase in volume gives the rate of blood flow into the arm or hand respectively. In the forearm, the muscles make up 85 per cent, the skin only 9 per cent of the volume, whereas in the hand and fingers about 15 per cent of the volume is muscle and 50 per cent skin; the rest is bone, the volume of which remains constant. Changes of blood flow into the forearm therefore reflect mainly vascular changes in the muscles, changes of blood flow in the hand and fingers, vascular changes in the skin. Barcroft and Edholm measured the changes in blood flow which occurred during post-haemorrhagic fainting mainly in students who volunteered for this experiment. Fainting was induced by putting rubber bands over the thighs so that the blood accumulated in the lower limbs, and then bleeding through venesection from an arm vein until fainting occurred. When this happened, there was a sudden and profound fall in arterial blood pressure associated with the following changes. In the hand the blood flow decreased and the skin became pale; i.e. vasoconstriction occurred in the cutaneous vessels. In the forearm the blood flow increased, i.e. vasodilatation occurred in the muscles. Since the vasodilatation was absent in patients with upper limb

inhibition of vasoconstrictor tone because nerve anaesthesia failed to release the spasm. Therefore it must have resulted from activation of sympathetic vasodilator fibres. The results have been confirmed and extended to the lower extremities by Fatherree and Allen (1938). Instead of using a hot air bath for warming the body, they immersed the body in hot water to meet a criticism made against the hot air bath method by Uprus, Gaylor and Carmichael (1936).

Lewis and Pickering were unable to display vasodilator, sympathetic fibres to the hand and fingers of healthy persons, because, in these, removal of the vasoconstrictor tone already raises the temperature of the skin so close to that of the blood entering it that activation of vasodilator impulses has little chance of displaying itself in a further rise of temperature. Grant and Holling (1937/8) showed that this does not hold true for the forearm and leg. This difference is attributed to the distribution of arterio-venous anastomoses which are present in the extremities of the limbs but not in the forearms and legs. In order to display in these parts activation of sympathetic vasodilators to the skin, such intense warming of the body is necessary that, when prolonged, it is not well tolerated. Warming was accomplished by immersing two, or better three, limbs in water maintained at 45° or 46° C. and covering the rest of the body, leaving exposed only the face and the limb under observation. This led to profuse sweating and, as evidenced by the flushing of the skin and the great rise in its temperature, to intense vasodilatation. The experimental proof that this vasodilatation resulted from activation of sympathetic vasodilators was the same as in the experiments of Lewis and Pickering. There was the following additional observation which showed convincingly that it must have been due to 'activation' and not to 'removal' of nerve impulses. When the skin vessels were first dilated and a cutaneous nerve then blocked whilst the intense heating was maintained, the area that became anaesthetic ceased to sweat, paled and cooled.

There is, however, this question: are we actually dealing with vasodilator fibres or is the vasodilatation in these experiments the result of activation of the cholinergic sympathetic fibres to the sweat glands? We cannot give a definite answer, but Grant

Sympathetic vasodilators in species other than man. We only need to discuss this innervation for the skeletal muscles. Proof for the presence of such fibres to the viscera rests on a few observations (Bunch, 1899; Bulbring and Burn, 1936) which are anything but convincing (Folkow, Frost and Uvnas, 1948; Eliasson, Folkow, Lindgren and Uvnas, 1951), and the skin lacks sympa-

been performed to find the conditions in which they are activated. There are also some observations suggesting a sympathetic vasodilator supply to the skin of the external ear (Bulbring and Burn, 1936), but the evidence is not sufficient and the observations have not been confirmed (Folkow, Frost, Haeger and Uvnas, 1949).

On the other hand, there is ample and convincing evidence

wild hare is not so strong, and there are apparently none present in monkeys (Macacus rhesus) nor in the tame or wild rabbit (Burn, 1932; Schneider, 1934; Bulbring and Burn, 1934/5; 1936; 1936/7, Rosenblueth and Cannon, 1935; Folkow and Uvnäs, 1948; 1950; Folkow, Haeger and Uvnäs, 1948; Eliasson, Folkow, Lindgren and Uvnas, 1951).

Bulbring and Burn have put forward the attractive hypothesis that the different development of sympathetic vasodilators to muscle is related to the power of endurance or capacity for muscular exertion of the animal. The monkey is easily fatigued; the rabbit is capable of running for a relatively short distance only; it can dash to the nearest rabbit hole. The cat can probably exert itself for a longer time than the rabbit, but both animals are greatly inferior to the dog which can run for miles without fatigue. The wild hare resembles the dog in this respect, and we may include man in this scheme as an animal with great power of endurance and a good sympathetic vasodilator supply to the muscles (Burn, 1938).

sympathectomies, it must have been mediated via the sympathetic. In order to decide whether it was the result of inhibition of vasoconstrictor tone or of excitation of sympathetic vasodilator fibres, Barcroft and Edholm compared the blood flow, during fainting, in a normal arm with that in an arm 'acutely sympathectomized' by nerve block: the vasodilatation was greater in the normal arm. Thus excitation of vasodilator sympathetic fibres must have occurred. Further, this vasodilatation is the actual cause of the profound and sudden fall in arterial blood pressure, because during the fall the cardiac output does not decrease. What happens when the haemorrhage reaches the critical stage is a sympathetic discharge which leads to vasoconstriction in the skin, but to a vasodilatation in the muscles which is so extensive that we bleed, so to speak, into our own vessels and thereby faint.

A similar excitation of the sympathetic vasodilators may occur under a number of circumstances, such as in sudden strong emotions, or in severe pain in which the subject is liable to fainting attacks. It is interesting to speculate with Barcroft and Edholm on how far this vasodilatation can be regarded as a protective mechanism which leaves the subject's circulation after recovery from the faint in a more stable condition than before, and if, in the special case of haemorrhage, it prevents a subject from bleeding outside by storing the blood in the muscles.

There is, as yet, no evidence that excitation of these sympathetic vasodilators to the muscles is involved in the normal physiological regulation of vasomotor tone. In this connection, it is worth while to emphasize that the vessels of the human skeletal muscles probably also receive vasoconstrictor fibres and possess vasoconstrictor tone (Barcroft, Bonnar, Edholm and Effron, 1943). It would be of interest to know if in man—as seems likely to be the case in animals—increase and decrease of vasoconstrictor tone is the sole mechanism of normal regulation of vasomotor tone, and if excitation of the sympathetic vasodilators is reserved for certain conditions of stress including severe muscular exertion, or if there is, in addition, a continuous interplay between the state of excitation in both nerve fibres.

Sympathetic vasodilators in species other than man. We only need to discuss this innervation for the skeletal muscles. Proof for the presence of such fibres to the viscera rests on a few observations (Bunch, 1899; Bulbring and Burn, 1936) which are anything but convincing (Folkow, Frost and Uvnas, 1948; Eliasson, Folkow, Lindgren and Uvnas, 1951), and the skin lacks sympathetic vasodilators except in certain special regions such as the

vated. There are also some observations suggesting a sympathetic vasodilator supply to the skin of the external ear (Bulbring and Burn, 1936), but the evidence is not sufficient and the observations have not been confirmed (Folkow, Frost, Haeger and Uvnas, 1949).

On the other hand, there is ample and convincing evidence for the presence of cholinergic sympathetic dilator fibres to the vessels of the limb muscles in dogs and cats. The evidence for the presence of sympathetic vasodilators to the limb muscles of the wild hare is not so strong, and there are apparently none present in monkeys (Macacus rhesus) nor in the tame or wild rabbit (Burn, 1932; Schneider, 1934; Bulbring and Burn, 1934/5; 1936; 1936/7; Rosenblueth and Cannon, 1935; Folkow and Uvnas, 1948, 1950; Folkow, Haeger and Uvnas, 1948; Eliasson, Folkow, Lindgren and Uvnas, 1951).

Bulbring and Burn have put forward the attractive hypothesis that the different development of sympathetic vasodilators to muscle is related to the power of endurance or

cat can probably exert itself for a longer time than the rabbit, but both animals are greatly inferior to the dog which can run for miles without fatigue. The wild hare resembles the dog in this respect, and we may include man in this scheme as an animal with great power of endurance and a good sympathetic vasodilator supply to the muscles (Burn, 1938).

If we accept this hypothesis, we should expect activation of these fibres during muscular exertion, but not in the normal regulation of vascular tone. Muscular exertion is one of the examples given by Cannon in which a widespread sympathetic discharge takes place. Since the centre for this discharge is in the hypothalamus, we should therefore expect that these fibres can be activated by hypothalamic stimulation.

The non-participation of the sympathetic vasodilators in depressor reflexes but their activation by hypothalamic stimulation was strikingly demonstrated by Folkow, Uvnäs and their co-workers in beautifully designed cross-circulation experiments on cats. The arrangement was such that the hind limbs of a cat in which a depressor reflex was elicited were perfused from another donor cat. By giving dibenamine to the donor, the vasoconstrictors of the perfused hind limbs could be eliminated without inactivating the cholinergic sympathetic vasodilators. In this condition, stimulation of the sympathetic abdominal chain no longer elicited a constrictor response in the perfused limbs, but a vasodilatation. But when depressor reflexes were elicited either by raising the arterial blood pressure or stimulating a sensory nerve, the vessels of the perfused limb muscles did not dilate. That is, the vasodilatation usually occurring in depressor reflexes is fully accounted for by inhibition of vasoconstrictor tone.

On the other hand, Eliasson, Folkow, Lindgren and Uvnas (1951) found that stimulation of certain points in the hypothalamus produced a vasodilatation in the hind limbs which persisted after dibenamine. The regions from which this effect was produced were not identical with those which produced vasoconstriction in skin and intestine. Fig. 6 gives the points that, when stimulated, gave muscle vasodilatation, they are between the area -5 to -9; vasoconstriction in skin and intestine was obtained in the area between -2 and -9. The muscle vasodilatation on hypothalamic stimulation was associated with other widespread sympathetic effects, which illustrates that activation of the sympathetic vasodilators to the muscles is an integral part of the reaction pattern of a general sympathetic discharge. that is, in Cannon's so-called emergency reaction. If

the emergency leads to a strenuous muscular exertion, the useful purpose of the discharge is more or less obvious. The ensuing dilatation of the muscles, reinforced by the local vasodilator metabolites, ensures a sufficient blood flow to the working muscle and, by reducing the peripheral resistance, keeps the

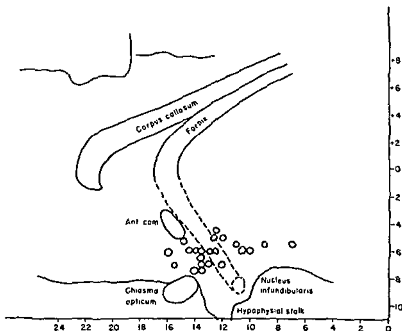


FIG. 6 Median section of hypothalamus showing points yielding muscular vasodilatation Anterior and vertical Horsley-Clarke co-ordinates are marked (after Eliasson, Folkow, Lindgren and Uvnäs, 1951)

arterial blood pressure within physiological ranges in spite of the enormous increase in cardiac output. The same discharge, however, unaccompanied by muscular exertion may, at least in man, precipitate a vasomotor collapse and fainting, as shown in the experiments of Barcroft and Edholm.

In times of emergency there is also an output of adrenaline. We must therefore inquire how the circulating adrenaline and noradrenaline affect the efficient working of the sympathetic vasodilators. Burn and Bulbring have shown that adrenaline

does not oppose but actually increases the vasodilatation, because adrenaline apparently sensitizes the muscle vessels to the dilator action of acetylcholine.

Tonic Activity in the Sympathetic System

There is no continuous sympathetic discharge to the sweat glands, to the intestinal muscles, the pilomotors or the suprarenal medulla. However, a large fraction of the preganglionic sympathetic outflow is constantly discharging nerve impulses. This so-called *tonic* activity is responsible for a number of sustained bodily functions, such as the vasomotor tone, whereby an adequate blood pressure is maintained. There is also some tonic sympathetic innervation to the heart. The tonic sympathetic innervation to the head region of the body is best illustrated by the effects which follow section of the cervical sympathetic: there is a certain degree of miosis, a sinking in of the eye (enophthalmus) and drooping of the upper lid, giving the eye a sleepy appearance. The tonic activity in the sympathetic forms a kind of background on which the local reflexes act. Cannon uses the following simile: 'the sympathetic is like the loud and soft pedals modulating all the tones together; the parasympathetic like the separate keys'.

Summary

If we contrast the functions of the two sub-divisions of the autonomic nervous system in the way we have—i.e. looking upon the parasympathetic as the efferent pathway for visceral reflexes where the viscera are localized, the sympathetic as the efferent pathway where the viscera are widespread and, in addition, realize that the sympatho-adrenal system can be activated as a whole and that tonic activity in both systems, but particularly in the sympathetic, forms a background on which the reflexes act—we obtain a better understanding of the function of the autonomic nervous system than by stressing the antagonism of the action of the two systems which has often been over-emphasized in the past.

REFERENCES

ASHER, L. (1909). *Z. Biol.* **52**, 298.

BACH, L. M. N. (1945/46). *Amer. J. Physiol.* **145**, 474.

BACQ, Z. M., BROUHA, L. and HEYMANS, C. (1932). *C.R. Soc. Biol., Paris*, **111**, 152.

BACQ, Z. M., BROUHA, L and HEYMANS, C. (1933a). *C.R. Soc. Biol., Paris*, **112**, 208.

BACQ, Z. M, BROUHA, L. and HEYMANS, C. (1933b). *C.R. Soc. Biol., Paris*, **114**, 154.

BACQ, Z. M., BROUHA, L and HEYMANS, C. (1934). *Arch. int. Pharmacodyn* **48**, 429.

BARCROFT, H., BONNAR, W. McK., EDHOLM, O. G. and EFFRON, A. S. (1943). *J Physiol.* **102**, 21.

BARCROFT, H. and EDHOLM, O. G. (1945). *J. Physiol.* **104**, 161.

339.

BERNTHAL, T., MOTLEY, H. E., SCHWIND, F. J. and WEEKS, W. F. (1945). *Amer J. Physiol.* **143**, 220.

BISHOP, G H, HEINBECKER, P. and O'LEARY, J L. (1933). *Amer. J. Physiol.* **106**, 647.

BOYD, J. D.

BROWN, G.

BRUCKE, F.

BULBRING, E.

BULBRING, E and BURN, J. H. (1936a). *J. Physiol.* **86**, 61.

BÜLBRING, E and BURN, J. H (1936b). *J Physiol.* **87**, 254.

BULBRING, E and BURN, J. H (1936/37) *J Physiol.* **88**, 341.

CELANDER, O and FOLKOW, B (1951). *Acta physiol. scand* **23**, 64.

CLARK, A J (1927). *Comparative Physiology of the heart.* Cambridge University Press.

DALE, H. H (1913) *J. Physiol* **46**, 291.

DALE, H H (1914) *J Pharmacol* **6**, 147.

DALE, H H. (1933). *J Physiol* **80**, 10 P.

DASTRE, A. and MORAT, J P. (1884). *Système nerveux vasomoteur.* Masson, Paris

DOLE, V P and MORISON, R. S. (1940). *Amer. J. Physiol.* **130**, 304.

ELIASSON, S, FOLKOW, B., LINDGREN, P. and UVNÄS, B (1951). *Acta physiol. scand.* **23**, 333

EULER, U. S. VON and GADDUM, J. H. (1931). *J. Physiol.* 73, 54.
EVANS, L. (1939). *Recent advances in physiology*. 6th edit. Churchill Ltd., London.
EVANS, L. and MATSUOKA, Y. (1915). *J. Physiol.* 49, 378.
FATHERREE, T. J. and ALLEN, E. V. (1938). *Arch. int. Med.* 62, 1015.
FELDBERG, W. (1950). *Arch. exp. Path. Pharmak.* 212, 64.
FOFANOW, L. L. and TSCHALUSSOW, M. A. (1913). *Pflüg. Arch. ges. Physiol* 151, 543.
FOLKOW, B., FROST, J., HAEGER, K. and UVNAS, B. (1948). *Acta physiol. scand* 15, 421
FOLKOW, B., FROST, J., HAEGER, K. and UVNAS, B. (1949). *Acta physiol. scand.* 17, 195.
FOLKOW, B, FROST, J. and UVNAS, B. (1948). *Acta physiol. scand.* 15, 412.
FOLKOW, B, HAEGER, K. and UVNAS, B. (1948). *Acta physiol. scand.* 15, 401.
FOLKOW, B., STROM, G. and UVNAS, B. (1950). *Acta physiol. scand.* 21, 145.
FOLKOW, B. and UVNAS, B. (1948a). *Acta physiol. scand.* 15, 365.
FOLKOW, B. and UVNAS, B. (1948b). *Acta physiol. scand.* 15, 389.
FOLKOW, B and UVNAS, B. (1950). *Acta physiol. scand.* 20, 329.
FREEMAN, N. E. and ROSENBLUETH, A. (1931) *Amer J. Physiol.* 98, 454.
GOLLWITZER-MEIER, KL., KROETZ, Chr. and KRUGER, E. (1938). *Pflug. Arch. ges. Physiol.* 240, 263.
GRANT, R T. and HOLLING, H. E. (1938) *Clin. Research*, 3, 273.
JARISCH, A (1925) *J. Physiol.* 60, 419.
LEWIS, T. and PICKERING, G. W. (1932). *Heart*, 16, 33.
MAREY, E.-J. (1863) *Physiologie médicale de la circulation du sang*, p. 206. A. Delahaye, Paris.
MARTIN, E. G and MENDELHALL, W. L (1915). *Amer. J Physiol* 38, 98.
MISLAVSKY, A. and FOFANOW, L. (1908). (Personal communication to W. M. Bayliss, quoted in the paper of 1908b)
NICKERSON, M. (1949). *J. Pharmacol. Part II. Pharmacol Rev.* 95, 27
PICK, J. and SHEEHAN, D (1946) *J. Anat., London.* 80, 12
PINKSTON, J. O, PARTINGTON, P. F and ROSENBLUETH, A. (1936). *Amer. J. Physiol.* 115, 711.
RANSON, S. W and MAGOUN, H. W (1939) *Ergebn Physiol* 41, 56.
ROSENBLUETH, A. and CANNON, W B. (1934) *Amer J Physiol.* 108, 599
SCHNEIDER, D. (1934). *Arch. exp. Path Pharmak* 176, 111
SKOOG, T (1947). *Lancet*, 2, 457
THOMAS, C. B and BROOKS, C. M. (1935) *Amer J Physiol* 113, 130P.
TOURNADE, A. and MALMÉJAC, J (1933) *C R Soc Biol, Paris.* 112, 679.
TSCHALUSSOW, M A. (1913) *Pflug Arch ges. Physiol.* 151, 523.
UPRUS, V, GAYLOR, J B and CARMICHAEL, E A. (1936) *Clin. Sci.* 2, 301
WRETE, M. (1935) *Morph. Jb* 75, 229
WYBAUW, L (1938). *Arch int Physiol* 46, 293

COMPLETE LIST OF LECTURES

** Lectures included in the published volumes*

1951–52

*The Scientific Approach to Medical Research
PROFESSOR E. ADRIAN, O.M., M D, F.R.C.P, P.R.S (*Cambridge*)

The Philosophy of Science
PROFESSOR H. DINGLE, D.SC., A.R.C S. (*London, University College*)

Studies by X-ray Analysis, Electron Microscopy and Supporting Techniques of the Structure of Biological Macromolecules and the Tissues formed from them
PROFESSOR W T. ASTBURY, M.A., SC.D., F.R.S. (*Leeds*)

*Tissue Responses to Injury
PROFESSOR G R CAMERON, M.B., B.S., D.SC., F R C P., F.R.S. (*London, University College Hospital Medical School*)

*Studies of Normal and Pathological Physiology of the Kidney
D. VAN SLYKE, M D. (*New York, U S.A.*)

Biochemical Function of Vitamin B_1
PROFESSOR R. A. PETERS, M.D, F.R S. (*Oxford*)

Nutritional Science and the Feeding of Populations
PROFESSOR S. J COWELL, M B, F R C.P. (*London, University College Hospital Medical School*)

*Nutritional Assessment of the Individual
PROFESSOR J YUDKIN, M.A., M.D, PH D. (*London, Queen Elizabeth College*)

Folic Acid and Vitamin B_{12} in the Metabolism of Micro-organisms
D. D. WOODS, M.A., PH D. (*Oxford*)

Aspects of Bacteriostasis
PROFESSOR B. C. J. G. KNIGHT, D.SC. (*Reading*)

*Some Aspects of Antibacterial Immunity
A. A. MILES, M.A., M.B., F.R.C.P.
(*National Institute for Medical Research, Mill Hill*)

*Antibodies and Immunity to Virus Infection
PROFESSOR A. W. DOWNIE, M.D., D.SC. (*Liverpool*)

Current Knowledge on Nature of Viruses
F. C BAWDEN, M.A., F.R.S. (*Rothamsted Experimental Station*)

*Viruses as the Causes of Diseases
PROFESSOR S. P. BEDSON, M.D., D.SC., F.R.C.P., F.R.S.
(*London Hospital Medical College*)

*Adrenal Hormones and ACTH
PROFESSOR F. G. YOUNG, M A., D.SC., F.R I.C., F.R.S. (*Cambridge*)

Chemical Aspect of Anti-biotics
SIR ROBERT ROBINSON, O.M., D SC, LL.D., F.R.S. (*Oxford*)

*Growth of the Human at the Time of Adolescence
J. M. TANNER, M D. (*London, St. Thomas's Hospital Medical School*)

Bone Growth and Repair
PROFESSOR W D. NEWCOMB, M.D, F R.C P
(*London, St Mary's Hospital Medical School*)

The Rate and Materials of Blood Formation
PROFESSOR L J. WITTS, D.M, F R.C P. (*Oxford*)

Alimentary Factors in Blood Formation
PROFESSOR L. J. WITTS, D M., F R C P (*Oxford*)

*Blood Coagulation in Theory and Practice
R. G MACFARLANE, M D. (*Oxford*)

The Psychology of Personality
PROFESSOR SIR CYRIL BURT, M A, D SC (*London, University College*)

*On the Organization of Cortical Mechanisms
PROFESSOR SIR GEOFFREY JEFFERSON, M S, F.R C.P, F R.C S, F.R S.
(*Manchester*)

Involuntary Nervous System (*2 lectures*)
PROFESSOR G. L. BROWN, C.B.E., M SC, F.R S.
(*London, University College*)

*The Physiological Basis of Visual Sensation
L. C. THOMSON, PH.D. (*London, Institute of Ophthalmology*)

*The Contribution of Science to the Practice of Health in the First Quarter of the Twentieth Century
PROFESSOR J. M. MACKINTOSH, M.D., F.R.C.P., LL.D.
(*London School of Hygiene and Tropical Medicine*)

Methods of Research in Social Medicine
ALICE M. STEWART, M.A., M.D., F.R.C.P. (*Oxford*)

Environmental Factors in Work
PROFESSOR G. P. CROWDEN, O.B.E., T.D., D.SC.
(*London School of Hygiene and Tropical Medicine*)

Methods of Research in Industrial Medicine
DONALD HUNTER, M.D., F.R.C.P. (*London Hospital Medical College*)

*Recent Light on Mammary Function
PROFESSOR H. D. KAY, C.B.E., D.SC., F.R.S. (*Reading*)

*Research on Ageing
PROFESSOR E. C. DODDS, M.D., F.R.C.P., F.R.S.
(*London, Middlesex Hospital Medical School*)

Carcinogenesis
PROFESSOR A. HADDOW, M.D., D.SC., PH.D.
(*London, Institute of Cancer Research*)

Physical Techniques in the Medical Applications of Ionizing Radiations
PROFESSOR W. V. MAYNEORD, D.SC.
(*London, Institute of Cancer Research*)

*Biological Effects of Radiation
J. F. LOUTIT, D.M., M.R.C.P. (*M.R.C., Harwell*)

*The Effects of Haemorrhage on the Cardiovascular System in Man
O. G. EDHOLM, B.SC., M.B., B.S.
(*National Institute for Medical Research, Mill Hill*)

*Lability of Blood Pressure
PROFESSOR M. L. ROSENHEIM, M.A., M.D., F.R.C.P.
(*London, University College Hospital Medical School*)

Cardiac Output in Man (*film*)
PROFESSOR J. M. MCMICHAEL, M.D., F.R.C.P.
(*Postgraduate Medical School of London*)

1952–53

The Influence of Language on Medicine
PROFESSOR J. Z. YOUNG, M.A., F.R.S. (*London, University College*)

The Nature and Place of Thinking in Medicine
PROFESSOR SIR FREDERIC BARTLETT, C.B.E., M.A., F.R.S. (*Cambridge*)

*The Methodology of Clinical Science
PROFESSOR SIR JAMES SPENCE, M.C., M.D., F.R.C.P. (*Newcastle*)

The Statistical Approach
PROFESSOR BRADFORD HILL, C.B.E., D.SC., PH.D. (*London School of Hygiene and Tropical Medicine*)

Control of Respiration
PROFESSOR C. G. DOUGLAS, C.M.G., M.C., M.A., M.D., F.R.S. (*Oxford*)

Life at High Altitudes
PROFESSOR SIR B. H. C. MATTHEWS, C.B.E., M.A., SC.D., F.R.S. (*Cambridge*)

*The Physiological Effects of Gravity
GROUP CAPTAIN W. K. STEWART, C.B.E., A.F.C., R.A.F. (*Farnborough*)

The Kinetics of Rapid Chemical and Biological Reactions
PROFESSOR F. J. W. ROUGHTON, M.A., PH.D., F.R.S. (*Cambridge*)

Metal Catalysis and Intracellular Respiration
PROFESSOR D. KEILIN, F.R.S. (*Cambridge*)

Biophysical Studies of Connective Tissue
PROFESSOR J. T. RANDALL, D.SC., F.R.S. (*London, King's College*)

*The Functional Significance of Connective Tissue
A. H. T. ROBB-SMITH, M.A., M.D., F.R.C.P. (*Oxford*)

*Silicosis
PROFESSOR E. J. KING, PH.D., D.SC. (*Postgraduate Medical School of London*)

Pulmonary Asbestosis
PROFESSOR MATTHEW STEWART, C.B.E., F.R.C.P. (*Leeds*)

Organ Culture in Biological and Medical Research
HONOR B. FELL, F.R.S. (*Cambridge*)

*The Life History of the Malaria Parasite
PROFESSOR P. C. C. GARNHAM, M.D., D.SC., D.P.H.
(*London School of Hygiene and Tropical Medicine*)

*Human Haemoglobins
J. C. WHITE, M B , CH.B. (*Postgraduate Medical School of London*)

Uses and Applications of Human Blood Plasma Fractions
SIR ALAN DRURY, C B E , M D., F.R S. (*Cambridge*)

*Preservation of Living Cells at Low Temperatures
A. S. PARKES, M.A., SC.D., F.R.S.
(*National Institute for Medical Research, Mill Hill*)

*The Life-Span of Red Blood-cells
P. L. MOLLISON, B.A., M.D. (*Postgraduate Medical School of London*)

*Biochemical Genetics
A. NEUBERGER, PH.D., M.D., F.R.S.
(*National Institute for Medical Research, Mill Hill*)

*Chromatography in the Study of Amino-Acid Metabolism
C. E. DENT, B.SC., PH.D., M.D., F.R.I C., M R.C.P.
(*London, University College Hospital Medical School*)

*Selective Toxicity (*2 lectures*)
PROFESSOR ADRIAN ALBERT, D SC., PH.D., F.R.I.C.
(*Australian National University*)

*Recent Progress in Antibiotics
PROFESSOR SIR ALEXANDER FLEMING, M.B , B.S., F.R.C.S., F.R.C.P., F.R S. (*London, St. Mary's Hospital Medical School*)

*Virus Adaptability in Relation to Human Disease
PROFESSOR WILSON SMITH, M D., F.R.S.
(*London, University College Hospital Medical School*)

*Carcinogenesis
P. R. PEACOCK, M.B , B S , F.R.F.P.S G. (*Glasgow*)

*The Metabolism of the Adrenocortical Hormones
PROFESSOR G. F. MARRIAN, D.SC., F.R.S. (*Edinburgh*)

The Pituitary Gland and Reproduction
S. J FOLLEY, D.SC., PH.D., F.R.S. (*Reading*)

*The Physiology of Parturition
PROFESSOR A. ST. G. HUGGETT, D.SC , PH.D.
(*London, St. Mary's Hospital Medical School*)

*The Physiology of the Autonomic Nervous System
W. S. FELDBERG, M.D., F.R.S.
(National Institute for Medical Research, Mill Hill)

*The Principles of Ganglionic Block
W. D. M. PATON, M.A., B M., B.CH.
(London, University College Hospital Medical School)

*Cholinesterases and Anti-Cholinesterases
PROFESSOR R. H. S. THOMPSON, D.M.
(London, Guy's Hospital Medical School)

Certain Aspects of Nerve Repair
H. J. SEDDON, M.A., D.M., F.R.C.S. *(London, Institute of Orthopaedics)*

Development of the Heart in Relation to Congenital Heart Disease
PROFESSOR J. D. BOYD, M.A., M.SC., M.D. *(Cambridge)*

Dynamics of the Heart Beat
PROFESSOR A. HEMINGWAY, M.SC., M.B., CH.B. *(Leeds)*

The Natural History of Essential Hypertension
PROFESSOR G. W PICKERING, M.A., M.B., F.R.C.P.
(London, St. Mary's Hospital Medical School)

Causes of Hypotension in Man
PROFESSOR E P SHARPEY-SCHAFER, M.B., B.CHIR., F.R.C.P.
(London, St. Thomas's Hospital Medical School)

PLATE I

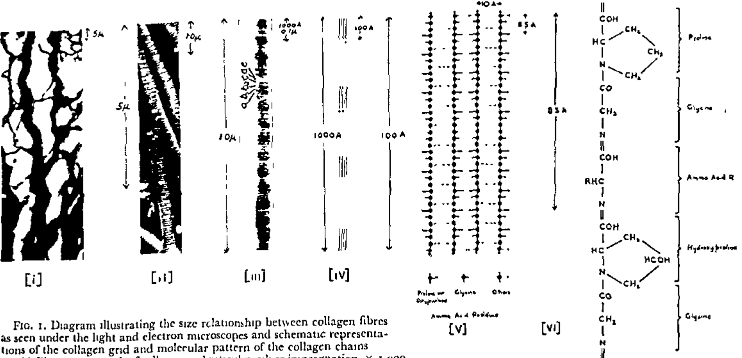


FIG. 1. Diagram illustrating the size relationship between collagen fibres as seen under the light and electron microscopes and schematic representations of the collagen grid and molecular pattern of the collagen chains

(i) Photomicrograph of collagen and reticulin, silver impregnation. × 1,000
(ii) Electron micrograph of collagen fibres shadowed with chromium × 14,000
(iii) Electron micrograph of collagen fibre stained with phosphotungstic acid. × 100,000. Reprinted by permission, from Gross, J. and Schmitt, F. O., 'Structure of human skin collagen as studied with the electron microscope'. *J. exper. Med* 88, 555 (1948).
(iv) Tenfold reduction of the collagen grid × 1,000,000
(v) Schema for the collagen grid. × 10,000,000 (modified from Astbury, W. T., *Journal Int. Soc Leath. Chem.* 24, 69 (1940)).
(vi) Schema for the molecular pattern of collagen chains. × 100,000,000

Plate II

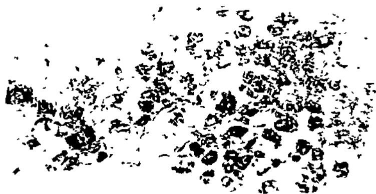


Fig 1 Silicotic nodulation. X-ray and section through part of the lung of a flint worker



FIG 2 Reticulation X-ray and section through part of the lung of a worker in shale and coal dust.

PLATE IV

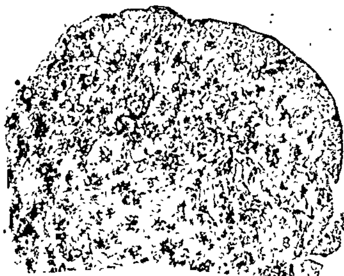


Fig. 3. Mixed nodulation and reticulation section through the lung of a worker in sandstone and shale

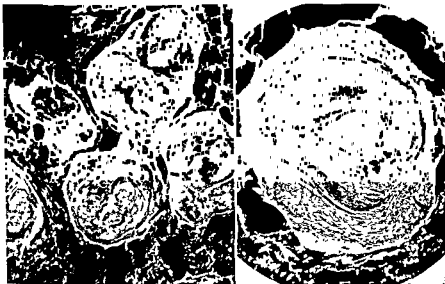


Fig 4 Human silicotic nodule low and high power magnification of sections of the lung of a flint worker

PLATE V

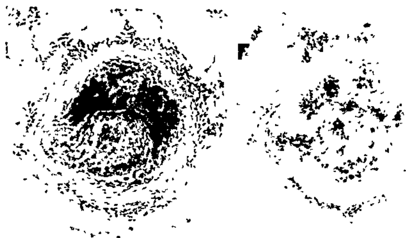


FIG 5 A human silicotic nodule: high power magnification of a single nodule Silverstain to show pattern of collagen fibrosis ('ball of twine' appearance), and micro-incineration to show pattern of siliceous dust (note 'core', vacant space, and 'halo') × 40

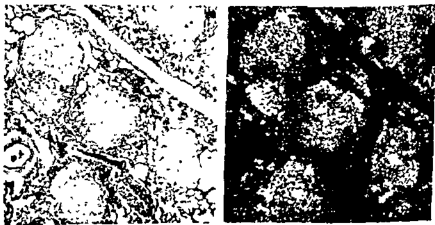


FIG 6. Quartz, experimental silicotic nodules low power magnitude of a section of lung from a rat which received powdered quartz Haematoxylin stain and micro-incineration × 24

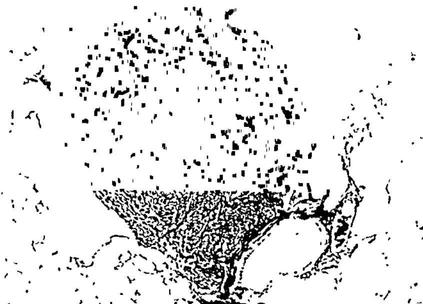


FIG. 7. Quartz, experimental silicotic nodule silver stain to show pattern of collagenous fibrosis in a single large nodule from the lung of a rat dusted with quartz for 400 days. × 75



FIG 8 Experimental silicotic fibrosis high power magnification of a section of a silicotic nodule to show intimate structure of collagen fibres Silver stain × 98

Fig 9 Kaolin section of the lung of a rat treated with South Wales kaolin Loose reticulinosis; minimum (grade 1) fibrosis despite presence of much dust. Silver stain.

Fig 10 Felspar: rat lung 500 days after 75 mg of felspar dust Focal accumulation of dust cells with some loose reticulin, grade 1 fibrosis Few dust cells in some alveoli. × 98

FIG. 11. Horneblende: rat lung 300 days after injection of 100 mg of horneblende schist, showing reticulinosis (grade 1 minimal) Silver stain × 30

FIG 12 Mica (sericite) focal accumulations of dust, mild reticulinosis and thickening of alveolar walls Haematoxylin and eosin stain × 60

PLATE IX

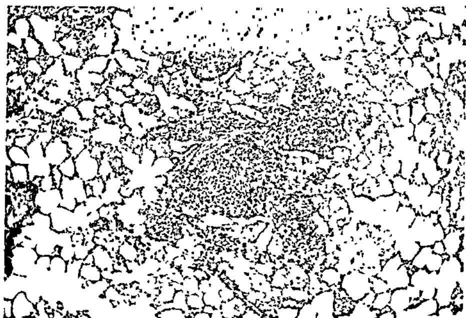


FIG. 13. Acid-treated mica (sericite): nodules of reticulin fibrosis (grade 3). Haematoxylin and eosin × 60

. . . 51 per cent mica, . . . Only a few fibres of . . . very dense accumulation of . . . r stain, and micro-incineration × 135

Fig. 15. Shale (8 per cent quartz, 77 per cent mica, 9 per cent kaolin). focal lesion showing mild fibrosis (grade 1). Similar to Fig 14

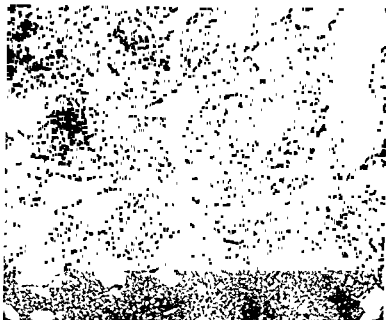


FIG 16 Quartz confluent collagenous fibrosis (grade 5) Large silicotic nodules in a rat's lungs 360 days Silver stain × 30

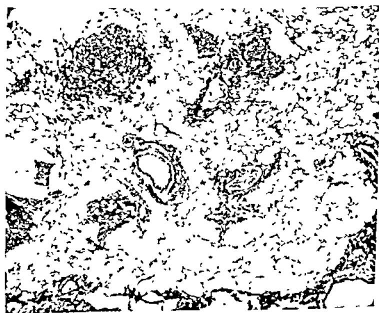


FIG 17 Quartz plus 1 per cent aluminium hydroxide scattered accumulations of dust in nodules of reticulin (grade 2 fibrosis). 360 days. Silver stain × 30



FIG. 18 Quartz· crowded nodules of densely packed collagen (grade 4 and 5 fibrosis) Rat dusted with powdered quartz for 400 days Silver stain × 30



FIG. 19 Quartz plus 2 per cent aluminium reticulin nodules (grades 1 and 2 fibrosis) 400 days' dusting with a mixture of powdered quartz with 2 per cent of powdered metallic aluminium Silver stain × 30

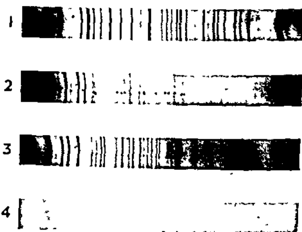
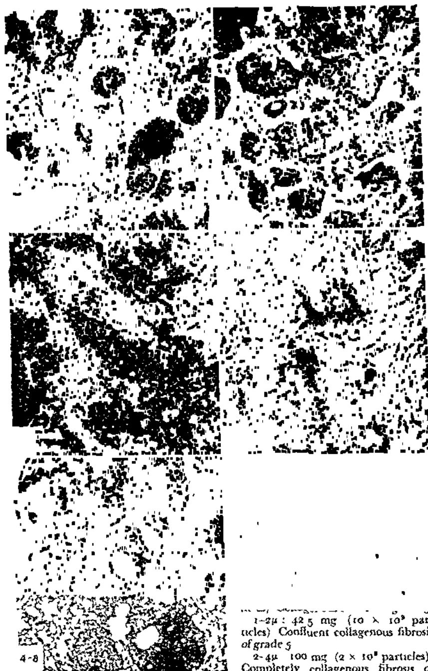


Fig. 20. X-ray diffraction patterns and electromicrographs of (1) quartz, (2) tridymite, (3) cristobalite and (4) fused (amorphous) silica

PLATE XIV



$1-2\mu$: 42.5 mg. (10×10^9 particles). Confluent collagenous fibrosis of grade 5

$2-4\mu$: 100 mg. (2×10^9 particles). Completely collagenous fibrosis of grade 5

$4-8\mu$: 199 mg. (0.5×10^9 particles). Almost collagenous nodules, partly acellular with ordinary nuclear stain. Grade 3 fibrosis early. $\times 25$

Plate XV

Key to identification of amino-acids shown on Plates XV–XXI

1, cysteic acid (from cystine); 2, aspartic acid, 3, glutamic acid, 4, phosphoethanolamine, 5, serine, 6, taurine, 7, glycine, 8, threonine; 9, alanine, 10, glutamine; 11, citrulline, 12, tyrosine; 13, tryptophan; 14, phenylalanine, 15, 'cat-spot'; 16, leucine and/or isoleucine, 17, valine; 18, α-amino-n-butyric acid, 19, β-amino-isobutyric acid, 20, histidine, 21, γ-aminobutyric acid; 22, proline, 23, 1- and 3-methylhistidines, 24, arginine; 25, lysine and ornithine, 26, protein (unidentified), 27, chloride

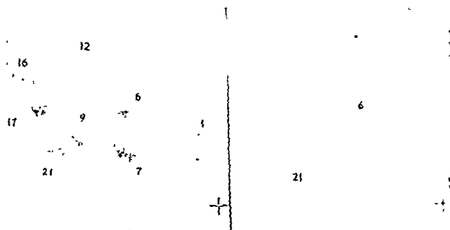


Fig 3 Photographs of chromatograms illustrating the 'copper trick' The same mixture has been analysed in both cases with the same two solvents in the same cabinet The paper on the right had however been first lightly dusted with basic copper carbonate in a line from right to left starting from the pencilled cross where the mixture had been placed In the first solvent the amino-acids track through the copper compound and the 'α' ones form complexes while the 'non-α's' do not Only the latter therefore appear as spots in their usual places on the final chromatogram.

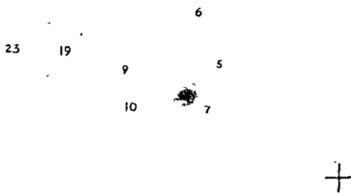


FIG. 5 Photograph of chromatogram obtained from normal human urine. The volume of urine containing 250 µg. of total N has been taken for the analysis in this and all subsequent chromatograms of urine

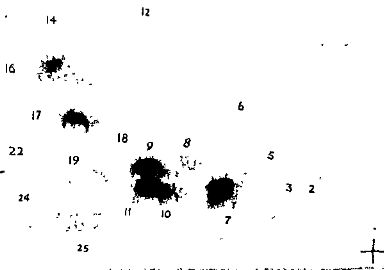


FIG 6 Photograph of chromatogram obtained from 625 µl. of normal human plasma ultrafiltrate

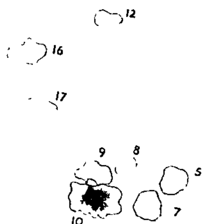


FIG 7 Photograph of chromatogram obtained from 625 µl of normal human cerebro-spinal fluid

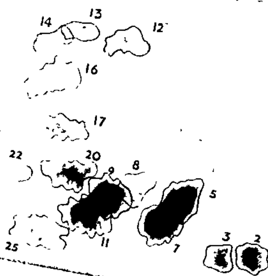


FIG 8 Photograph of chromatogram obtained from 125 µl. of human ...at

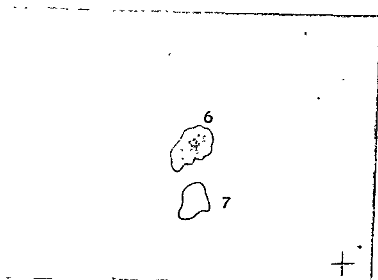


Fig. 10. Photograph of chromatogram obtained from mouse urine.

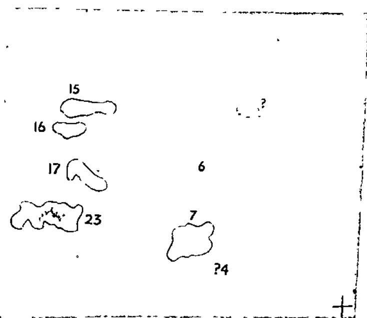


Fig. 11 Photograph of chromatogram obtained from the urine of a domestic cat.

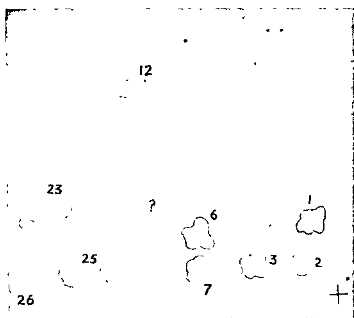


FIG. 12 Photograph of chromatogram obtained from the urine of the blotched genet



FIG. 13 Photograph of chromatogram obtained from the urine of a case of Fanconi syndrome (adult type)

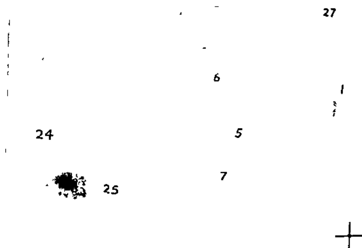


Fig 14 Photograph of chromatogram obtained from the urine of a case of *stone-forming cystinuria.*

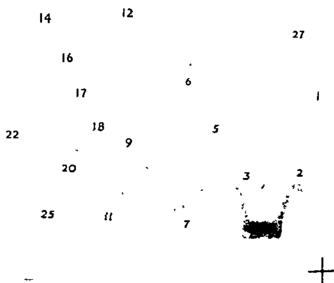


Fig 15 Photograph of chromatogram obtained from the urine of a case *of Wilson's disease*

PLATE XXI



FIG 16 Photograph of chromatogram obtained from the urine of a case of 'Hart's syndrome' The numbers refer to colour strengths on an arbitrary scale

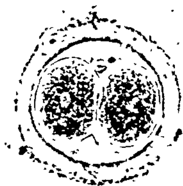


FIG. 2 Fertilized rabbit egg Obtained before segmentation, and treated with glycerol at 40° C Cooled to −79 C, left at −190° C for 20 hours, and thawed at +40° C Glycerol removed at +40° C Cultured for 15 hours, during which time segmentation occurred (From Dr Smith's material) × 250

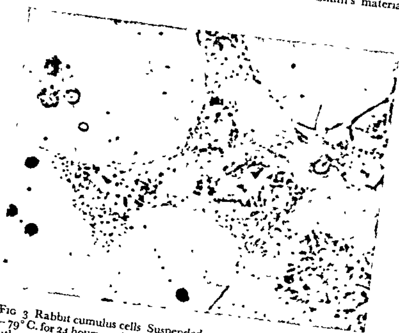


FIG 3 Rabbit cumulus cells Suspended in serum Cooled to −79° C, left at −79° C. for 24 hours, and thawed at +40° C. Cultured for 18 hours (From Smith, 1952b) × 500

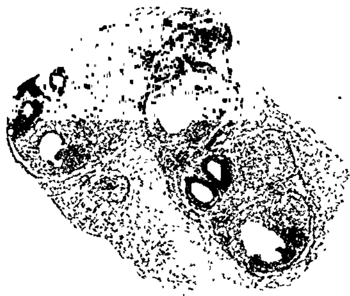


FIG 4 Rat ovarian tissue 28 days after grafting The tissue had been chopped in 15 per cent glycerol-saline, cooled slowly to − 79° C, and stored at − 190° C. for 117 days before grafting (From Parkes and Smith, 1953.) × 23



FIG 5 Rat ovarian tissue 13 weeks after grafting The tissue had been chopped in 15 per cent glycerol, cooled slowly to − 79° C. and stored at − 79° C for 83 days before grafting (From Parkes and Smith, 1953) × 23

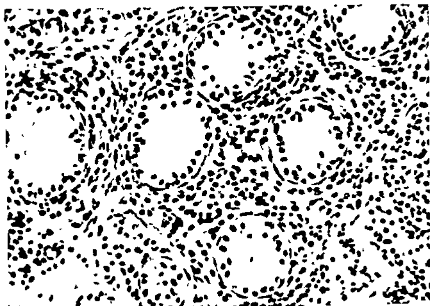


Fig. 6. Rat testis tissue 111 days after grafting The testes of infantile rats had been teased out and soaked in 15 per cent glycerol-saline, cooled slowly to − 79° C. and stored at − 79° C for 7 days before grafting to another individual. × 240

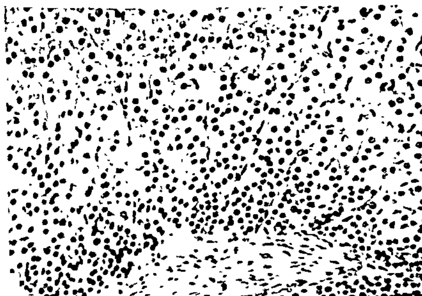


Fig 7. Rat adrenal tissue 42 days after grafting Four half-adrenals had been 15 per cent glycerol, cooled slowly to − 79° C. and kept at − 79° C. ur before grafting. × 240

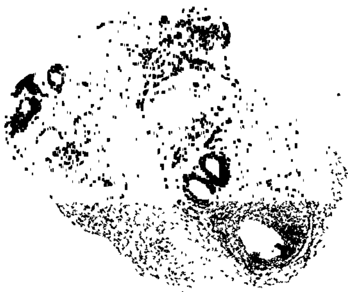


FIG 4 Rat ovarian tissue 28 days after grafting The tissue had been chopped in 15 per cent glycerol-saline, cooled slowly to − 79° C, and stored at − 190° C, for 117 days before grafting (From Parkes and Smith, 1953.) × 23

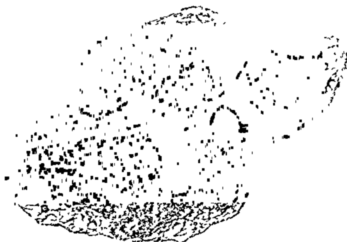


FIG 5. Rat ovarian tissue 13 weeks after grafting. The tissue had been chopped in 15 per cent glycerol, cooled slowly to − 79° C and stored at − 79° C. for 83 days before grafting (From Parkes and Smith, 1953) × 23



polarizer
-axis lies

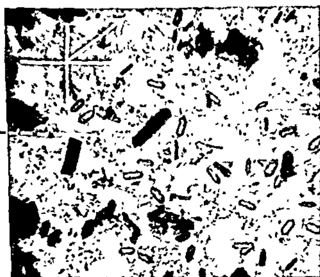


FIG 2. Crystals of human reduced haemoglobin between crossed-nicols, with first order gypsum plate introduced to show positive sign of birefringence. × 75

The crystals in Figs 1 and 2 were separated from pH 6 7 phosphate buffer, r/2 4·75

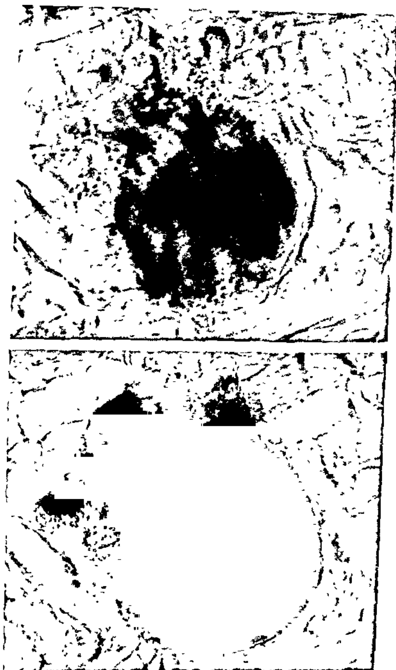


FIG 8 Amoebae suspended in 5 per cent glycerol in rain-water, and cooled under the microscope. × 370 (From Smith, Polge and Smiles, 1951.)
(*above*) Medium crystallized at − 10° C.
(*below*) Intracellular crystallization at − 15° C.

Plate XXVII

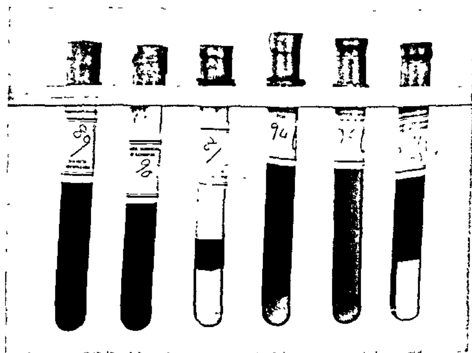


Fig 6 Solubility of human reduced haemoglobins in pH 6 7 phosphate buffer at r/2 4 75

89.	Normal maternal sample
90	Full-term infant, cord sample
81, 95	Sickle-cell anaemia
94.	Sickle-cell anaemia, ? microdrepanocytic disease
91	Sickle trait

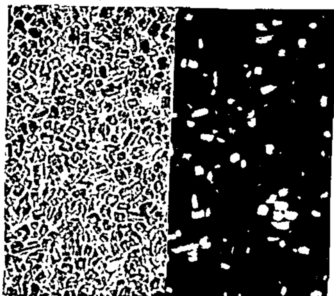


FIG 10 A typical congenital haemolytic anaemia. Erythrocytes in buffered, isotonic sodium dithionite (Itano, H. A., and Pauling, L. (1949) *Blood*, 4, 66). *Left* Normal illumination. *Right* Polarized light, showing birefringent crystals of reduced haemoglobin distorting the cells × 500.



FIG. 11 Sickle-cell anaemia, preparation as in Fig 5 *Left* Normal illumination *Right*. Polarized light, showing birefringence of sickle cells. × 500